



TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

SUVI ALTIO
**PIRKANMAALLA 2004–2008 TAPAHTUNEIDEN LIIKENNE-
ONNETTOMUUKSIEN ANALYSOINTIA BAYESVERKKOJEN
AVULLA**

Diplomityö

Tarkastaja: Yliopistonlehtori

Esko Turunen

Tarkastaja ja aihe hyväksytty

Luonnontieteiden ja ympäristötekniikan
tiedekunnan tiedekuntaneuvoston

kokouksessa 15.8.2012

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO

Teknis-luonnontieteellinen koulutusohjelma

ALTIO, SUVI: Pirkanmaalla 2004–2008 tapahtuneiden liikenneonnettomuuksien analysointia bayesverkkojen avulla

Diplomityö, 53 sivua, 3 liitesivua

Joulukuu 2012

Pääaine: Matematiikka

Tarkastaja: Yliopistonlehtori Esko Turunen

Avainsanat: bayesverkot, riippuvuusmallit, B-course, liikenneonnettomuudet

Tässä työssä testataan Complex Systems Computation Group -tutkijaryhmän kehittämän B-course-ohjelmiston riippuvuuksien mallinnusmenetelmän soveltuvuutta liikenneonnettomuusdatan tutkimiseen. Ohjelmiston datan analysointi perustuu bayesverkkoihin, jotka ovat todennäköisyysmalleja. Riippuvuuksien mallinnusmenetelmä etsii mallin, joka parhaiten kuvaa aineiston muuttujien välisiä riippuvuussuhteita. B-course-ohjelmistossa on Java Playground -sovellus, jolla riippuvuusmallia voidaan tutkia tarkemmin.

Tutkittavana aineistona on tietoja Pirkanmaalla vuosien 2004 – 2008 aikana tapahtuneista liikenneonnettomuuksista. Riippuvuusmallinnuksen testauksessa aineistosta tutkitaan muuttujien välisiä riippuvuussuhteita, niiden vahvuuksia sekä kausaalisuussuhteiden suuntaa ja laatua. Lisäksi testataan Java Playground -sovellusta tarkastelemalla lähemmin taajamissa ja taajamien ulkopuolella tapahtuneiden liikenneonnettomuuksien eroja, sekä eläinonnettomuuksia ja yli 65-vuotiaitten aiheuttamia liikenneonnettomuuksia.

Työn tuloksena voidaan todeta, että B-course-ohjelmiston riippuvuuksien mallinnusmenetelmä soveltuu hyvin riippuvuussuhteiden löytämiseen käytetyn liikenneonnettomuusdatan tyypisille aineistoille. Paljon muuttujia sisältävä aineisto on syytä jakaa pienempiin osiin tuloksena saatavan riippuvuusmallin tarkastelun helpottamiseksi. Tuloksina saatujen riippuvuus- ja kausaalisuusmallien muuttujien välisten suhteiden laatuun on hyvä suhtautua kriittisesti. Erityisesti kausaalisuussuhteiden suunta ei vaikuttaisi aina olevan sen suuntainen kuin loogisin perustein pitäisi olla.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Science and Engineering

ALTIO, SUVI: Analysis of Traffic Accidents Happened in Pirkanmaa during 2004–2008 with Bayesian Networks

Master of Science Thesis, 53 pages, 3 Appendix pages

December 2012

Major: Mathematics

Examiner: Associate Professor Esko Turunen

Keywords: Bayesian networks, dependency models, B-course, traffic accidents

In this thesis, the dependency modeling method of B-course software is tested by studying the data of the traffic accidents. B-course software is developed by a research group Complex Systems Computation Group. The analysis made by the software is based on Bayesian networks, which are probability models. Dependency modeling method searches for the dependency model, which best defines the dependencies between the variables of the data. B-course software includes a Java Playground application that can be used to study the dependency model more closely.

The examined data consists of traffic accidents that have happened in Pirkanmaa during 2004 – 2008. In the testing of the dependency modeling the strength of dependencies between the variables, and the direction and the quality of the causalities between the variables in the data are examined. In addition, the Java Playground application is tested more closely by examining the traffic accidents that have happened inside and outside urban areas, accidents involving animals, and the accident that have been caused by over 65 year olds.

As a result, it can be said that the dependency modeling of the B-course software is well suited for finding dependencies in the type of data the traffic accidents represent. Data containing a lot of variables should be divided into smaller parts so that the resulting dependency model is easier to interpret. However, the resulting dependency and causality models should be met criticism. Especially the direction of causality does not always seem to be in the direction that can be logically concluded.

ALKUSANAT

Tämä diplomityö on tehty omalla vapaa-ajalla töiden ja nykyisten opintojeni ohessa Tampereen teknillisen yliopiston Matematiikan laitokselle. Työn aiheesta ja ohjauksesta haluan kiittää työni ohjaajaa yliopistonlehtori Esko Turusta.

Kiitän perhettäni tuesta, josta olen saanut nauttia koko opintojeni ajan. Haluan kiittää myös työkavereitani kuluneista vuosista ja kannustuksesta opintojeni suhteen. Kiitos myös ystävilleni, joiden seurassa olen kahvikupin ääressä tai lenkipolulla saanut vaihtaa monia ajatuksia. Lopuksi suuri kiitos Jussille ja Ronjalle päivieni piristämisestä, sekä erityiskiitos Jussille diplomityöni lukemisesta ja kommentoinnista.

Jyväskylässä, 5.11.2012

Suvi Altio

SISÄLLYS

| | |
|---|----|
| 1. Johdanto | 1 |
| 2. Riippuvuusmallit | 3 |
| 2.1 Riippuvuus ja Bayesin kaava | 3 |
| 2.2 Kausaalisuus | 4 |
| 2.3 Riippuvuuksien mallintaminen ja bayesverkot | 4 |
| 2.4 Naiivi kausaalisuusmalli | 5 |
| 2.5 Ei-niin-naiivi kausaalisuusmalli | 7 |
| 3. Aineiston käsittely ja mallin etsiminen B-course-ohjelmistolla | 10 |
| 3.1 Aineiston käsittely | 10 |
| 3.1.1 Puuttuvan tiedon käsittely | 10 |
| 3.1.2 Luokitteluasteikolliset muuttujat | 11 |
| 3.2 Mallin etsiminen | 11 |
| 3.2.1 Mallien lukumäärä | 11 |
| 3.2.2 Parhaan mallin etsiminen | 12 |
| 3.2.3 Riippuvuusmallien todennäköisyydet | 12 |
| 3.2.4 Esimerkki todennäköisyyden $P(D/M)$ laskemisesta | 15 |
| 4. Tiedot liikenneonnettomuuksista Pirkanmaalla vuosina 2004–2008 | 21 |
| 4.1 Tietoja kuljettajasta, keliolosuhteista ja onnettomuus- ajankohdasta | 21 |
| 4.2 Onnettomuuden tyyppi ja luokittelu | 23 |
| 4.3 Tietoja onnettomuuspaikasta | 24 |
| 5. Aineiston jakaminen pienempiin kokonaisuuksiin | 27 |
| 5.1 Muuttujakokonaisuuksien rajaamisprosessi | 27 |
| 5.1.1 Muuttujien karsiminen | 27 |
| 5.1.2 Muuttujien arvojen muokkaaminen | 28 |
| 5.1.3 Muuttujakokonaisuuksiin jakaminen | 28 |
| 5.2 Muuttujakokonaisuuksien riippuvuusmallit | 29 |
| 5.2.1 Muuttujakokonaisuuden 1 riippuvuusmalli | 30 |
| 5.2.2 Muuttujakokonaisuuden 2 riippuvuusmalli | 32 |
| 5.2.3 Muuttujakokonaisuuden 3 riippuvuusmalli | 32 |
| 5.2.4 Muuttujakokonaisuuden 4 riippuvuusmalli | 32 |
| 5.2.5 Muuttujakokonaisuuden 5 riippuvuusmalli | 34 |
| 5.2.6 Muuttujakokonaisuuden 6 riippuvuusmalli | 35 |
| 6. Tuloksia | 37 |
| 6.1 Taajamissa ja taajamien ulkopuolella tapahtuneiden liikenneonnettomuuksien erot Pirkanmaalla | 37 |
| 6.2 Eläinonnettomuudet Pirkanmaalla | 41 |

| | |
|---|----|
| 6.3 Yli 65-vuotiaiden aiheuttamat liikenneonnettomuudet | |
| Pirkanmaalla | 45 |
| 7. Yhteenveto | 50 |
| Lähteet | 53 |
| A.Liikenneonnettomuustyytit | 54 |

KUVAT

| | | |
|-----|---|----|
| 2.1 | Riippuvuusmallivaihtoehtoja muuttujille A , B ja C | 6 |
| 2.2 | Riippuvuusmallivaihtoehtoja toteuttamaan riippuvuuslauseita II.1–4. | 8 |
| 3.1 | B-course-ohjelman antama paras riippuvuusmalli esimerkin aineistolle. | 16 |
| 5.1 | Riippuvuusmallit muuttujakokonaisuuksille 1 ja 2. Erilaiset kaaret muuttujien välillä kuvaavat erilaisia muuttujien välisiä riippuvuussuhteita. | 31 |
| 5.2 | Riippuvuusmallit muuttujakokonaisuuksille 3 ja 4. Erilaiset kaaret muuttujien välillä kuvaavat erilaisia muuttujien välisiä riippuvuussuhteita. | 33 |
| 5.3 | Muuttujakokonaisuuden 5 riippuvuusmalli. Erilaiset kaaret muuttujien välillä kuvaavat erilaisia muuttujien välisiä riippuvuussuhteita. | 34 |
| 5.4 | Muuttujakokonaisuuden 6 riippuvuusmalli. Erilaiset kaaret muuttujien välillä kuvaavat erilaisia muuttujien välisiä riippuvuussuhteita. | 35 |
| 6.1 | Muuttujien <i>taaj-merk</i> , <i>nop-raj</i> , <i>nop-suuntaan</i> , <i>valoisuus</i> ja <i>tunti</i> riippuvuussuhteet toisiinsa sekä muuttujaan <i>onluokka</i> | 43 |
| 6.2 | Muuttujien <i>taaj-merk</i> , <i>nop-suuntaan</i> ja <i>saa</i> riippuvuussuhteet toisiinsa sekä muuttujaan <i>ika</i> | 48 |

TAULUKOT

| | | |
|------|--|----|
| 3.1 | Esimerkin aineisto. Arvot ovat keksittyjä eivätkä vastaa työssä tutkittavaa aineistoa. | 15 |
| 3.2 | Muuttujien <i>sukupuoli</i> ja <i>taajama</i> arvojen yhdistelmät, ja näiden yhdistelmien lukumäärät ($N_{ijk:t}$) aineistossa. | 17 |
| 3.3 | Konfiguraatiot muuttujan <i>viikonloppu</i> vanhemmille. | 18 |
| 3.4 | Muuttujien <i>taajama</i> , <i>alkoholi</i> ja <i>viikonloppu</i> arvojen yhdistelmät, ja näiden yhdistelmien lukumäärät ($N_{ijk:t}$) aineistossa. | 18 |
| 3.5 | Muuttujien <i>sukupuoli</i> ja <i>alkoholi</i> arvojen yhdistelmät, ja näiden yhdistelmien lukumäärät ($N_{ijk:t}$) aineistossa. | 19 |
| 4.1 | Muuttujien <i>saa</i> , <i>pinta</i> ja <i>valoisuus</i> arvojen selitykset. | 22 |
| 4.2 | Muuttujan <i>kp-luokka</i> arvoja vastaavat tien hoitoluokat. | 22 |
| 4.3 | Muuttujien <i>paallyste</i> ja <i>paal-lk</i> arvojen selitykset. | 23 |
| 4.4 | Muiden omaisuusvahinkojen selitykset muuttujalle <i>muu-vah</i> | 23 |
| 4.5 | Onnettomuusluokat muuttujalle <i>onluokka</i> | 23 |
| 4.6 | Asutustiheyden luokittelu muuttujan <i>asutus</i> arvoille. | 24 |
| 4.7 | Nopeusrajoituslajit, jotka vastaavat muuttujan <i>nop-laji</i> arvoja. | 24 |
| 4.8 | Muuttujan <i>onn-paikka</i> arvojen selitykset. | 25 |
| 4.9 | Muuttujien <i>junction-type</i> ja <i>solmun-tyyp</i> arvojen selitykset. | 26 |
| 4.10 | Muuttujien <i>muu-liit</i> ja <i>tasol-tyyppi</i> arvojen selitykset. <i>muu-liit</i> (muu liittymä) kuvaa liittymien erikoistapauksia ja <i>tasol-tyyppi</i> (liittymän luokka) kuvaa tavallisten tasoliittymien luokkia. | 26 |
| 4.11 | Muuttujan <i>rist-var</i> arvojen selitykset, jotka kuvaavat onnettomuuspaikan liikenteenohjausta. | 26 |
| 5.1 | Aineiston muuttujakokonaisuudet. | 29 |
| 6.1 | Muuttujakokonaisuuden 2 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan <i>taaj-merk</i> arvojen K ja E suhteen. Muuttujien arvojen selitykset ovat luvussa 4. | 38 |
| 6.2 | Muuttujakokonaisuuden 1 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan <i>taaj-merk</i> arvojen K ja E suhteen. Taulukosta puuttuvat muuttujat <i>muu vah</i> , <i>nop-vastaan</i> , <i>nak-460</i> ja <i>saa</i> , koska niiden jakaumat löytyvät taulukosta 6.1. Muuttujien arvojen selitykset ovat luvussa 4. | 39 |

- 6.3 Muuttujakokonaisuuden 3 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *taaj-merk* arvojen K ja E suhteen. Taulukosta puuttuu muuttujat *nak-150*, *muu vah*, *nop-vastaan*, *kp-luokka* ja *toim-luokka*, koska niiden jakaumat löytyvät taulukoista 6.1 ja 6.2. Muuttujan *onn-paikka* arvo 7 ei esiinny aineistossa kertaakaan. Muuttujien arvojen selitykset ovat luvussa 4. 40
- 6.4 Muuttujakokonaisuuden 4 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *onluokka* arvojen 10 – 12 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *onluokka* ei ole valittu arvoa. Muuttujien arvojen selitykset ovat luvussa 4. 42
- 6.5 Muuttujien *nop-suuntaan*, *nop-raj*, *valoisuus*, *tunti* ja *taaj-merk* arvojen todennäköisyysjakaumat muuttujan *onluokka* arvojen 10 – 12 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *onluokka* ei ole valittu arvoa. Muuttujien arvojen selitykset ovat luvussa 4. 44
- 6.6 Muuttujakokonaisuuden 4 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *ika* arvojen 65 – 93 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *ika* ei ole valittu arvoa. Muuttujien arvojen selitykset ovat luvussa 4. 46
- 6.7 Muuttujakokonaisuuden 5 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *ika* arvojen 65 – 93 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *ika* ei ole valittu arvoa. Taulukosta puuttuvat muuttujat *nak-150*, *muu vah* ja *nop-vastaan*, koska niiden jakaumat löytyvät taulukosta 6.6. Muuttujien arvojen selitykset ovat luvussa 4. 47
- 6.8 Muuttujien *nop-suuntaan*, *taaj-merk* ja *saa* arvojen todennäköisyysjakaumat muuttujan *ika* arvojen 65 – 93 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *ika* ei ole valittu arvoa. Muuttujien arvojen selitykset ovat luvussa 4. 49

LYHENTEET

CoSCo

Complex Systems Computation Group

HIIT

Helsinki Institute for Information Technology

MERKINNÄT

| | |
|--------------------|---|
| A | Muuttuja |
| B | Muuttuja |
| C | Muuttuja |
| D | Aineisto |
| E_i | Tapahtuma |
| F | Tapahtuma |
| K | Latentti muuttuja |
| L | Latentti muuttuja |
| M | Malli |
| M_i | Malli |
| n | Muuttujien lukumäärä |
| N' | Ekvivalentti otoskoko |
| N_{ij} | Niiden rivien lukumäärä datassa, joilla on j :s konfiguraatio i :n muuttujan vanhemmista |
| N_{ijk} | Niiden rivien lukumäärä datassa, joilla on k :s arvo i :nnelle muuttujalle, jolle on j :s konfiguraatio |
| p | Todennäköisyys |
| P | Todennäköisyys |
| P_i | Osatodennäköisyys, apumuuttuja aineiston todennäköisyyden laskemiseksi |
| q_i | i :n muuttujan vanhempien mahdollisten konfiguraatioiden lukumäärä |
| r_i | i :n muuttujan arvojen lukumäärä |
| S | Muuttujajoukko |
| t | Muuttuja |
| v_i | Satunnaismuuttujan arvo |
| x | Satunnaismuuttuja |
| X_i | Satunnaismuuttuja |
| Y | Solmu |
| θ | Parametri |
| Γ | Gammafunktio |
| η | Satunnaismuuttuja |
| π | Jeffreysin prior |
| $[I(\theta)]_{ij}$ | Fisherin informaatiomatriisi |

AINEISTON MUUTTUJAT

| | |
|----------------------|---|
| <i>ajor-lkm</i> | Ajoratojen lukumäärä onnettomuuspaikalla |
| <i>asutus</i> | Asukastiheys onnettomuuspaikan alueella |
| <i>ika</i> | Kuljettajan ikä |
| <i>junction-type</i> | Tieliikenneverkon solmun tyyppi |
| <i>kk</i> | Kuukausi, jolloin onnettomuus on tapahtunut |
| <i>kp-luokka</i> | Onnettomuuspaikan tien hoitoluokka |
| <i>kulj-alkoh</i> | Kuljettaja alkoholin vaikutuksen alaisena |
| <i>kuolleet</i> | Onnettomuudessa kuolleiden lukumäärä |
| <i>kvl</i> | Keskimääräinen vuorokausiliikenne |
| <i>lampotila</i> | Onnettomuusajankohdan lämpötila |
| <i>liik-valot</i> | Liikennevalojen toiminta |
| <i>loukkaant</i> | Onnettomuudessa loukkaantuneiden lukumäärä |
| <i>mo/mol</i> | Moottori(liikenne)tie |
| <i>muu-liit</i> | Muu liittymä |
| <i>muu vah</i> | Onnettomuuden muut omaisuusvahingot |
| <i>nak-150</i> | Näkemäprosentti 150 metrin etäisyydelle |
| <i>nak-300</i> | Näkemäprosentti 300 metrin etäisyydelle |
| <i>nak-460</i> | Näkemäprosentti 460 metrin etäisyydelle |
| <i>nop-laji</i> | Nopeusrajoituslaji |
| <i>nop-raj</i> | Onnettomuuspaikalla vallitseva nopeusrajoitus |
| <i>nop-suuntaan</i> | Onnettomuuden aiheuttaneen kulkuneuvon nopeus |
| <i>nop-vastaan</i> | Onnettomuuteen joutuneen vastaantulijan nopeus |
| <i>onluokka</i> | Onnettomuusluokka |
| <i>onn-paikka</i> | Onnettomuuspaikka |
| <i>ontyyppi</i> | Onnettomuustyyppi |
| <i>osallkm</i> | Onnettomuuteen joutuneiden osapuolten lukumäärä |
| <i>paallyste</i> | Onnettomuuspaikan tien päällyste |
| <i>paal-lk</i> | Tien päällysteluokka |
| <i>pinta</i> | Tien pinnan tila onnettomuuspaikalla |
| <i>promil</i> | Kuljettajan puhaltama promillelukema |
| <i>raskas-kvl</i> | Raskaiden kulkuneuvojen keskimääräinen vuorokausiliikenne |
| <i>rask-osall</i> | Raskas kulkuneuvo osallisena |
| <i>rist-var</i> | Risteyksen tyyppi |
| <i>saa</i> | Onnettomuusajankohdan säätila |
| <i>solmun-tyyp</i> | Tieliikenneverkon solmun tyyppi |

| | |
|---------------------|--|
| <i>sukup</i> | Kuljettajan sukupuoli |
| <i>suuntalkm</i> | Eri suuntiin kulkevien ajoratojen lukumäärä |
| <i>taaj-merk</i> | Onnettomuus tapahtunut taajamassa |
| <i>tasol-tyyppi</i> | Liittymäluokka |
| <i>tien-lev</i> | Onnettomuuspaikan tien leveys |
| <i>toim-luokka</i> | Tieluokka |
| <i>tunti</i> | Onnettomuuden kellonaika tasatunneiksi pyöristettynä |
| <i>valoisuus</i> | Onnettomuusajankohdan valoisuusaste |
| <i>vkpv</i> | Viikontähti, jolloin onnettomuus on tapahtunut |

1. JOHDANTO

Tiedonlouhinnalla (*engl. data mining*) tarkoitetaan menetelmiä, joiden avulla tietomassasta pyritään löytämään tietomassan käyttäjälle merkityksellisiä asioita. Nykyään tietoa kerätään lähes joka asiasta, ja tallennetut tietomassat voivat olla todella suuria. Tyypillisiä tiedonkeräysmenetelmiä ovat esimerkiksi yritysten asiakastietokannat, valtion ja kuntien hallinnolliset rekisterit ja teollisuusprosessien yhteydessä suoritettava mittaustulosten keruu. Yleensä suuret tietomassat koostuvat siis systemaattisesti kerätyistä tiedoista. Tiedonlouhintaan on kehitetty monia erilaisia tietokonemenetelmiä, joiden avulla suuria aineistoja pystytään tutkimaan.

Tutkittavat aineistot ovat yleensä todella suuria, ja monet tiedonlouhintaohjelmien laskenta-algoritmeista vievät paljon laskenta-aikaa. Lisäksi useimpien ohjelmien käyttäminen ja tulosten tulkinta vaatii käyttäjältä asiantuntemusta myös tiedonlouhinnasta eikä vain tutkittavan aineiston aiheesta. Tutkijaryhmä Complex Systems Computation Group (CoSCo) Helsingin yliopiston ja Aalto-yliopiston alaiselta Tietotekniikan tutkimuslaitokselta, Helsinki Institute for Information Technology (HIIT), on pyrkinyt löytämään ratkaisun tähän ongelmaan kehittäessään B-course-ohjelmiston.

B-course on internetissä käytettävä tiedonlouhintaohjelma, joka analysoi aineistoa bayesverkkojen avulla. Ohjelmistoon voi ladata oman taulukkomuotoisen datansa. B-course-ohjelmiston kapasiteetti on melko rajallinen, eikä se sovellu suurten aineistojen tutkimiseen. Käyttäjää opastetaan tiedoston lataamisessa ja tulosten tulkinassa. Tällä tavoin on pyritty siihen, että kuka tahansa voi käyttää B-coursea perehtymättä ohjelmiston perustana olevaan matemaattiseen teoriaan, ja analysointi voidaan suorittaa lyhyessä ajassa. Ohjelmistossa on kaksi bayesverkkoihin perustuvaa menetelmää datan analysointia varten. Toisella menetelmällä voidaan etsiä aineistosta riippuvuussuhteita, ja toisella menetelmällä voidaan tutkia valitun muuttujan arvojen luokittelua muiden muuttujien avulla.

Tässä työssä testataan B-course-ohjelmiston riippuvuussuhteiden mallinnusmenetelmän soveltuvuutta Pirkanmaalla vuosien 2004 – 2008 aikana tapahtuneiden liikenneonnettomuuksien analysointiin. Tutkittava data on osa aineistoa, jossa on koko Suomessa samana aikana tapahtuneiden liikenneonnettomuuksien tiedot. Koko Suomen kattavaa aineistoa on tutkittu muilla tiedonlouhintamenetelmillä Jyväskylän yliopistossa ja Tampereen teknillisessä yliopistossa [1, 2]. Tätä työtä varten aineis-

toa on rajattu pienemmäksi B-course-ohjelmiston rajallisen kapasiteetin takia. Työn tavoitteena ei ole löytää uusia tuloksia tutkitusta aineistosta, vaan testata millaisia riippuvuuksia B-course-ohjelmistolla aineistosta löytää. Alueellisesti rajattu aineisto poikkeaa monilta osin koko Suomen aineistosta, joten tämän työn tuloksia ei voi suoraan verrata muilla menetelmillä saatuihin tuloksiin. Aineistosta tutkitaan muuttujien välisten riippuvuussuhteiden vahvuuksia sekä kausaalisuussuhteiden suuntaa ja laatua. Lisäksi tarkastellaan B-course-ohjelmistoon kuuluvan Java Playground-sovelluksen avulla taajamissa ja taajamien ulkopuolella tapahtuneitten liikenneonnettomuuksien eroja, sekä eläinonnettomuuksia ja yli 65-vuotiaitten aiheuttamia liikenneonnettomuuksia.

Loppuosa työstä on jaoteltu seuraavasti. Toisessa luvussa esitellään Bayesin kaava ja riippuvuusmallit. Kolmannessa luvussa tutustaan B-course-ohjelman tapaan käsitellä aineistoa, sekä ohjelman taustalla olevaan matemaattiseen teoriaan. Tutkittava aineisto esitellään luvussa neljä. Luvussa viisi on esitetty aineiston jakaminen pienempiin muuttujakokonaisuuksien ja B-coursen antamat riippuvuusmallit näille kokonaisuuksille. Kuudennessa luvussa esitellään tarkempia tuloksia muuttaman muuttujan osalta. Lopuksi viimeisessä luvussa tehdään yhteenveto.

2. RIIPPUVUUSMALLIT

Työssä käytettävän B-course-ohjelmiston riippuvuuksien mallintaminen perustuu bayesiläiseen tilastotieteeseen. Tässä luvussa tutustutaan bayesiläisen tilastotieteen perustana olevaan Bayesin kaavaan, muuttujien väliseen riippuvuuteen sekä erilaisiin riippuvuusmalleihin, jotka tässä työssä ovat bayesverkkoja.

2.1 Riippuvuus ja Bayesin kaava

Klassisessa tilastotieteessä tilastollinen riippuvuus perustuu todennäköisyyksiin. Riippuvuuden matemaattinen tulkinta voidaan tehdä riippumattomuuden määritelmien avulla.

Määritelmä 2.1.1 (Riippumattomuus) *Satunnaismuuttujat X_1, X_2, \dots, X_n , ovat riippumattomia, jos*

$$P(X_1 = v_1, X_2 = v_2, \dots, X_n = v_n) = \prod_{i=1}^n P(X_i = v_i) \quad (2.1)$$

kaikille arvojoukoille $\{v_1, v_2, \dots, v_n\}$. [3]

Määritelmä 2.1.2 (Parittainen ehdollinen riippumattomuus) *Satunnaismuuttujat X_1 ja X_2 ovat ehdollisesti riippumattomia toisistaan, jos*

$$P(X_1 = v_1 | X_2 = v_2) = P(X_1 = v_1) \quad (2.2)$$

ja

$$P(X_2 = v_2 | X_1 = v_1) = P(X_2 = v_2) \quad (2.3)$$

kaikille arvoille v_1 ja v_2 . [3]

Muuttujat ovat keskenään *riippuvia*, jos määritelmän 2.1.1 ehto ei täyty. Kaksi muuttujaa on *parittain ehdollisesti riippuvia*, jos määritelmän 2.1.2 ehdot eivät täyty. Kaksi muuttujaa on parittain ehdollisesti riippuvia toisistaan, jos toisen arvon tietäminen auttaa toisen arvon arvaamisessa. Bayesiläisessä tilastotieteessä todennäköisyydet liittyvät uskomuksiin, ja siten bayesiläinen tulkinta riippuvuudelle on vapaampi kuin klassisen tilastotieteen tulkinta. Tapahtuman todennäköisyys määräytyy saatujen tietojen perusteella, ja saattaa siten muuttua, kun tietoa tulee lisää. Bayesiläisessä tilastotieteessä todennäköisyydet ovat aina ehdollisia toden-

näköisyyksiä ja perustuvat *Bayesin kaavaan*

$$P(E_i|F) = \frac{P(F|E_i)P(E_i)}{\sum_{j=1}^n P(F|E_j)P(E_j)} = \frac{P(F|E_i)P(E_i)}{P(F)}, \quad (2.4)$$

missä E_i :t ja F ovat tapahtumia, $P(E_i)$ on tapahtuman E_i prioritodennäköisyys, $P(E_i|F)$ on posterioritodennäköisyys, $P(F|E_i)$ on tapahtuman F ehdollinen todennäköisyys ehdolla E ja $P(F)$ on tapahtuman F prioritodennäköisyys. Prioritodennäköisyydet ovat todennäköisyyksiä, joiden määrittämisessä ei huomioida muiden tapahtumien vaikutusta. Posterioritodennäköisyys on tapahtuman E ehdollinen todennäköisyys ehdolla F , jonka suuruus halutaan selvittää, kun tapahtuma F on havaittu. [3]

B-course-ohjelma perustuu bayesiläiseen tilastotieteeseen, ja kaikki ohjelman malleissa esiintyvät riippuvuudet ovat parittaisia ehdollisia riippuvuuksia [4, 5]. Tästä lähtien tässä työssä riippuvuudella tarkoitetaan parillista ehdollista riippuvuutta.

2.2 Kausaalisuus

Kausaalinen riippuvuussuhde muuttujien välillä tarkoittaa sitä, että toinen muuttujista on syy ja toinen seuraus. Esimerkiksi muuttujien *vuodenaika* ja *hukkumiskuolemat* (henkilöä kuukaudessa) välinen suhde voidaan asioita yksinkertaistaen ymmärtää kausaalisen suhteen: muuttujaa *vuodenaika* voidaan ajatella syynä ja *hukkumiskuolemat*-muuttujaa seurauksena. Kesällä ihmiset liikkuvat enemmän vesillä päihtyneinä, joten myös hukkuneita on kesällä enemmän kuin talvella.

Joskus kausaalisen riippuvuussuhteen syynä saattaa olla latentti muuttuja. Se on muuttuja, joka ei esiinny mallissa, mutta jos se olisi mallissa mukana, sillä olisi syy-seuraussuhde niihin kahteen muuttujaan, joiden välillä mallissa on kausaalisuus-suhde. Näissä molemmissa suhteissa latenttimuuttuja olisi syy ja mallissa esiintyvät muuttujat seurauksia. Esimerkiksi muuttujien *hukkumiskuolemat* (henkilöä kuukaudessa) ja *jäätelönsyönti* (litraa kuukaudessa per henkilö) välillä esiintyy kausaalinen suhde, vaikka niiden välillä ei ole syy-seuraussuhdetta. Kausaalinen suhde kuitenkin selittyy latentilla muuttujalla *vuodenaika*: kesällä ihmiset syövät enemmän jäätelöä ja hukkumiskuolemat ovat yleisempiä kuin talvella.

2.3 Riippuvuuksien mallintaminen ja bayesverkot

Riippuvuuksien mallintamisessa etsitään mallia, joka kuvaa muuttujien välisiä todennäköisiä riippuvuuksia. Mallinnuksen tavoitteena on löytää kaikki aineiston muuttujien väliset riippuvuussuhteet. Mallin avulla voidaan tehdä päätelmiä muuttu-

jien välisistä kausaalisuussuhteista ja tutkittavan aineiston rakenteesta. Bayesiläisen tilastotieteen avulla voidaan vertailla eri malleja niiden todennäköisyyksien perusteella, ja valita malleista paras eli se, jonka todennäköisyys on suurin. [4, 5] Tämän todennäköisyyden laskeminen on esitelty luvussa 3.2.3.

Bayesverkot ovat verkkotodennäköisyysmalleja, jotka koostuvat solmuista ja niiden välisistä suunnatuista kaarista. Solmut vastaavat aineiston muuttujia ja solmujen väliset kaaret muuttujien välisiä ehdollisia riippuvuuksia. Bayesverkossa solmuja yhdistävät kaaret eivät saa muodostaa silmukoita. Silmukka tarkoittaa tilannetta, jossa kulkemalla kaarien suuntaisesti päästään takaisin solmuun, josta lähdettiin. [6]

Oletetaan, että on riippuvuusmalli, jossa on muuttujat A ja B sekä muuttujajoukko S (voi olla tyhjä), johon kuuluvat ne mallin muuttujat, joista tiedetään jotakin, lisäksi malli voi sisältää muuttujia, joista ei tiedetä mitään. Jos A :sta B :hen pääsee kulkemalla mallin kaaria pitkin, A ja B ovat riippuvia toisistaan. Jos A :sta ei pääse kulkemaan B :hen, ne eivät ole riippuvia. Kaaria pitkin kuljettaessa ei tarvitse huomioida suunnattujen kaarien suuntaa paitsi törmäyssolmujen kohdalla. [6]

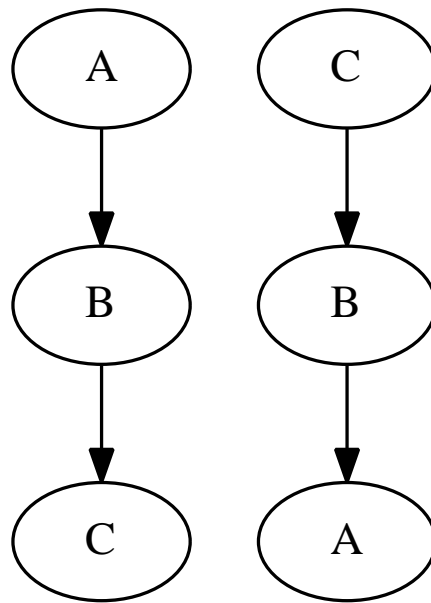
Törmäyssolmu on solmu, johon tullaan kulkemalla kaarta kaaren suunnan suuntaisesti (eteenpäin), ja josta poistuessa kuljetaan kaarta kaaren suuntaa vastaan (taaksepäin). Törmäyssolmun läpi kulkemiseen vaikuttaa se, kuuluuko törmäyssolmu joukkoon S vai ei. S :n ollessa tyhjä joukko riippuvuusmallissa voidaan kulkea kaaria pitkin eteen- ja taaksepäin kunhan ei kertaakaan vierailla samassa solmussa kahdesti tai päädytä törmäyssolmuun. Jos S ei ole tyhjä joukko, mallissa kulkeminen hankaloituu. Päätyminen sellaiseen S :n solmuun, joka ei ole törmäyssolmu, katkaisee reitin. Sen sijaan joukkoon S kuuluvien törmäyssolmujen läpi voidaan kulkea. Lisäksi sellaisten joukkoon S kuulumattomien törmäyssolmujen läpi kulkeminen onnistuu, joiden *lapsisolmu* kuuluu joukkoon S . Solmun Y lapsella tarkoitetaan sellaisia solmuja, jotka voidaan tavoittaa seuraamalla kaarta, joka lähtee Y :stä. [6]

2.4 Naiivi kausaalisuusmalli

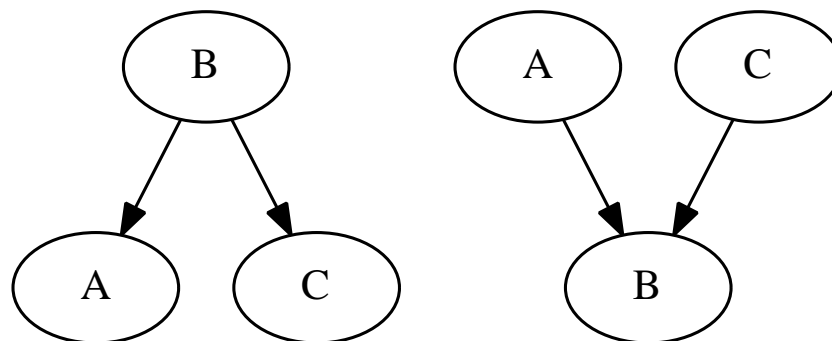
Kausaalisuuden päättely tilastollisista riippuvuuksista perustuu kausaalisuuden ja tilastollisten riippuvuuksien ominaisuuksiin. Yleisesti voidaan ajatella, että kaikki riippuvuussuhteet ovat kausaalisia. *Naiivissa kausaalisuusmallissa* oletetaan lisäksi, että kaikki mallissa esiintyvät muuttujien väliset riippuvuudet johtuvat mallin muuttujien välisistä kausaalisista suhteista. Tämä tarkoittaa siis sitä, että mikään mallin ulkopuolinen muuttuja ei voi aiheuttaa mallissa olevien muuttujien välisiä riippuvuussuhteita. Dataa analysoitaessa naiivin kausaalisuusmallin avulla mahdollisten latenttien muuttujien olemassa oloa ei tule huomioida, vaan riippuvuussuhteiden syyt tulee löytää mallista. [5]

Naiivin riippuvuusmallin muuttujien välisten suhteiden kausaalisuuden suuntaa ei voida aina tietää. Joskus kausaalisuuden suunta voidaan kuitenkin päätellä. Oletetaan, että on riippuvuusmalli, joka sisältää muuttujat A , B ja C sekä muuttujajoukon S (S voi olla tyhjä). Muuttujajoukko S on joukko, jonka muuttujista tiedetään jotain. Oletetaan sitten, että mallissa on voimassa **riippuvuuslauseet I**:

1. A ja B ovat aina riippuvia
2. B ja C ovat aina riippuvia
3. A ja C eivät ole riippuvia, kun on S , joka ei sisällä B :tä (eli B ei tunneta)



(a) Mahdollisuus 1. (b) Mahdollisuus 2.



(c) Mahdollisuus 3.

(d) Mahdollisuus 4.

Kuva 2.1: Riippuvuusmallivaihtoehtoja muuttujille A , B ja C .

Riippuvuuslauseen I.3 perusteella A ei voi olla suora seuraus C :stä, eikä C voi olla suora seuraus A , koska näissä tapauksissa A ja C olisivat aina riippuvia toisistaan.

Nyt on siis olemassa neljä eri konfiguraatiota muuttujille A , B ja C sekä niiden välisille suhteille, jotka on esitelty kuvissa 2.1(a)-2.1(d).

Mahdollisuus 1: A on B :n suora syy ja B on suora syy C :lle (kuva 2.1(a)).

Tässä tapauksessa A ja C olisivat aina riippuvia toisistaan, jos S on annettu, sillä S ei sisällä B :tä. B :stä jotain tietäminen olisi ainoa tapa estää riippuvuus A :n ja C :n välillä (B :n kautta). Myös tämä tapaus on riippuvuuslauseen I.3 vastainen.

Mahdollisuus 2: C on suora syy B :lle ja B on suora syy A :lle (kuva 2.1(b)).

Tässä tapauksessa tilanne on sama kuin mahdollisuudessa 2, riippuvuus A :n ja C :n välillä on vain päinvastainen. Tämänkin tapaus on siis riippuvuuslauseen I.3 vastainen.

Mahdollisuus 3: B on yhteinen suora syy A :lle ja C :lle (kuva 2.1(c)).

Tässä tapauksessa, jos B :tä ei tiedetä, A ja C vaikuttaisivat olevan riippuvia toisistaan, sillä tuntematon yhteinen syy saa seuraukset näyttämään riippuvilta. Tämä on kuitenkin riippuvuuslauseen I.3 vastainen tilanne, koska A ja C eivät saa olla riippuvia toisistaan.

Mahdollisuus 4: B on yhteinen suora seuraus A :sta ja C :stä (kuva 2.1(d)).

Nyt B on törmäyssolmu, joten se estää riippuvuuden A :n ja C :n välillä. Tämä on ainoa malli, joka on kaikkien riippuvuuslauseiden mukainen. [5]

2.5 Ei-niin-naiivi kausaalisuusmalli

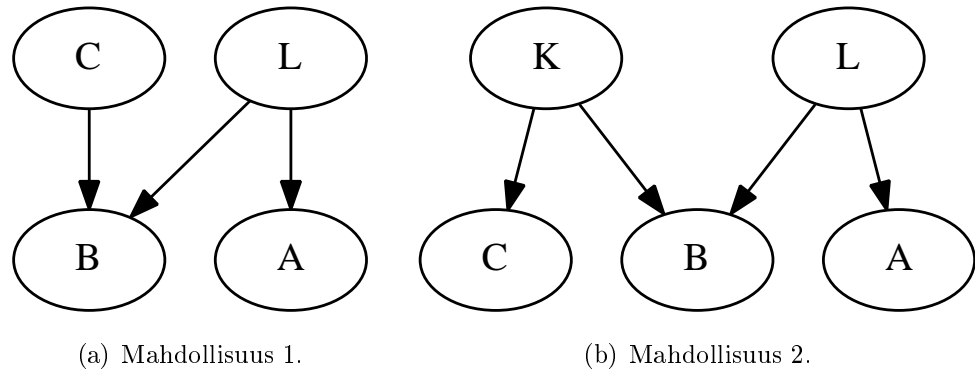
Useimmiten kaikkien riippuvuussuhteiden syynä voi olla latentti muuttuja. B-course-ohjelmistossa latentteja muuttujia sisältävissä malleissa rajoitutaan malleihin, joissa jokainen latentti muuttuja on tasan kahden tarkasteltavan muuttujan vanhempi, ja latenteilla muuttujilla ei ole vanhempia. Tällaisessa tapauksessa voidaan päätellä jotakin tarkasteltavien muuttujien kausaalisista suhteista, ja näitä malleja kutsutaan B-course-ohjelmistossa *ei-niin-naiiveiksi kausaalisuusmalleiksi*. Tällainen rajoitus on hyväksyttävää, koska on mahdollista todistaa, että sopivilla oletuksilla kaikki kausaalisuusmallit, joissa on latentteja muuttujia, voidaan esittää tämän latenttien muuttujien suhteen rajoitetun luokan malleina. [5]

Muuttujien välisten kausaalisuussuhteiden selittäminen latenteilla muuttujilla ei ole aina kovin yksinkertaista. Joskus muut tarkasteltavan kausaalisuussuhteen lähellä olevat riippuvuudet voivat auttaa poistamaan mahdollisuuden, että A aiheuttaa B :n, vaikka A ja B vaikuttavat olevan aina riippuvia. Tällaisessa tapauksessa on olemassa muuttujajoukko S ja kolmas muuttuja C , joka on riippuva A :sta mutta riippumaton B :stä. Jos A on suora syy B :hen, riippuvuus A :n ja C :n välillä tekisi aina myös B :n ja C :n riippuviksi toisistaan, mikä on vastoin oletusta, jonka mukaan B ja C ovat riippumattomia toisistaan, kun on olemassa joukko S . Jäljelle jää kaksi mahdollisuutta: joko B aiheuttaa A :n tai A :lle ja B :lle on yhteinen latentti syy. [5]

Joskus malli mahdollistaa sulkemaan pois sen vaihtoehdon, että A :lla ja B :llä on yhteinen latentti syy, jolloin jäljelle jää vaihtoehto, että B on A :n suora syy. Muutetaan hieman edellä esitettyä tilannetta, ja tarkastellaan seuraavat ehdot täyttävää mallia. Oletetaan, että mallissa on muiden riippuvuuksien joukossa seuraavanlaiset **riippuvuuslauseiden II** mukaiset riippuvuudet (B ei kuulu joukkoon S):

1. A ja B ovat aina riippuvia toisistaan
2. B ja C ovat aina riippuvia toisistaan
3. A ja C ovat riippuvia, kun on joukko S
4. A ja C eivät ole riippuvia, kun on joukko S ja tunnetaan B

Oletetaan lisäksi edellä esitettyjä riippuvuuslauseita II hyödyntäen, että B ei ole C :n suora syy. Nyt on mahdotonta, että yhteinen latentti syy voisi aiheuttaa riippuvuuden A :n ja B :n välillä, koska tässä tapauksessa, jos B tunnetaan, syntyy A :n ja C :n välille riippuvuussuhde. Tällainen tilanne ei toteuta riippuvuussuhdetta II.4. Tämä voidaan todeta tarkastelemalla kahta mahdollista tilannetta, joissa A :lla ja B :llä on yhteinen latentti syy L (kuvat 2.2(a), 2.2(b)).



Kuva 2.2: Riippuvuusmallivaihtoehtoja toteuttamaan riippuvuuslauseita II.1–4.

Mahdollisuus 1: C aiheuttaa B :n ja latentti muuttuja L aiheuttaa A :n ja B :n (kuva 2.2(a)).

Tässä tapauksessa A ja C vaikuttaisivat riippuvilta, jos B tunnetaan. B on tässä tapauksessa törmäyssolmu, ja jos se tunnetaan, sen läpi voidaan kulkea. Nyt C :stä voidaan kulkea B :n ja L :n läpi A :han. Tämä tapaus on riippuvuuslauseen II.4 vastainen.

Mahdollisuus 2: Latentti muuttuja K aiheuttaa sekä C :n että B :n, ja latentti muuttuja L aiheuttaa sekä A :n että B :n (kuva 2.2(b)).

Tässä tapauksessa A ja C vaikuttaisivat riippuvilta, jos B tunnetaan. B on tässäkin tapauksessa törmäyssolmu, ja jos se tunnetaan, sen läpi voidaan kulkea. Nyt C :stä

voidaan kulkea K :n, B :n ja L :n läpi A :han. Tämä tapaus on riippuvuuslauseen II.4 vastainen.

Jotta malli on riippuvuuslauseiden mukainen, siinä ei ole latentteja muuttujia. Ainoa vaihtoehto on siis se, että B on A :n suora syy. [7]

3. AINEISTON KÄSITTELY JA MALLIN ETSIMINEN B-COURSE-OHJELMISTOLLA

B-course-ohjelmisto analysoi vain luokitteluasteikollisia muuttujia. Tutkittava aineisto voi kuitenkin sisältää myös muuttujia, joiden arvot ovat diskreettejä, jatkuvia, järjestysasteikollisia. Ohjelmisto muuttaa näiden muuttujien arvot luokitteluasteikollisiksi. Parhaan riippuvuusmallin etsimisessä ohjelmisto vertailee eri mallien todennäköisyyksiä keskenään ja pyrkii löytämään suuresta joukosta todennäköisimmän mallin. Tässä luvussa esitellään parhaiten aineistoa vastaavan riippuvuusmallin etsintäprosessin matemaattista taustaa.

3.1 Aineiston käsittely

3.1.1 Puuttuvan tiedon käsittely

Puuttuvan tiedon käsittelyn yksinkertaisin menetelmä on jättää analyysistä pois datarivit, joiden tiedoissa on puutteita, ja suorittaa analyysi vain jäljellä olevien datarivien perusteella. Toisena vaihtoehtona on puuttuvan tiedon imputointi, eli puuttuvien tietojen arvaaminen tai muulla tavalla korvaaminen, ja tämän jälkeen käyttää koko aineistoa analyysin suorittamisessa. B-course-ohjelma suorittaa puuttuvan tiedon käsittelyn yhdistäen edellä esiteltyjä tapoja. Ohjelma pyrkii poistamaan analysoitavasta aineistosta vain ne osat datariveistä, joissa ei ole tietoja, ei siis kokonaisia rivejä. Tämän jälkeen ohjelma suorittaa analyysin jäljellä olevalle aineistolle. [5, 8]

B-course-ohjelman tavoitteena on löytää todennäköisin riippuvuusmalli, joten ohjelman pitää pystyä vertailemaan eri mallien todennäköisyyksiä. Tämä vertailu perustuu ohjelman kykyyn laskea aineiston D todennäköisyys perustuen ohjelman antamaan malliin M . [4, 5] Tämän todennäköisyyden $P(D|M)$ laskeminen esitellään luvussa 3.2.3. Kaava todennäköisyyden $P(D|M)$ laskemiselle on paikkansapitävä vain täydelle datalle, mutta B-course käyttää samaa kaavaa myös puutteellisille aineistoille.

3.1.2 Luokitteluasteikolliset muuttujat

B-Course-ohjelma edellyttää muuttujilta luokitteluasteikollisuutta, jotta se pystyy muodostamaan muuttujien välisen riippuvuusmallin. Jos muuttujat eivät ole luokitteluasteikollisia, ohjelma muuttaa ne sellaisiksi. Jatkuvien numeeristen muuttujien tiedot muutetaan luokitteluasteikollisiksi diskretoinnin avulla. Kun jatkuvat ja diskreetit numeeriset muuttujat sekä järjestysasteikolliset muuttujat muutetaan luokitteluasteikollisiksi, tieto muuttujien arvojen järjestyksestä katoaa. [4, 5]

Muuttujien ollessa luokitteluasteikollisia aineiston analyysissä päästään eroon mahdollisista jakaumaoletuksista ja muuttujien välisistä lineaarisista riippuvuussuhteista. Kun muuttujiin liittyviä oletuksia on vähemmän, voidaan löytää malli, joka sisältää epälineaarisia ja mahdollisesti monimutkaisempia riippuvuussuhteita muuttujien välille. [5]

Jatkuva-arvoisten muuttujien diskretisoinnin toteutus vaikuttaa tuloksena saatavaan riippuvuusmalliin. Mitä vähemmän luokkia diskretoinnin avulla muodostetulla luokitteluasteikolla on, sitä enemmän riippuvuussuhteita löytyy. Myös diskretisoinnissa käytettävien jakopisteiden valinnalla on vaikutusta lopputulokseen. Varmistaakseen halutunlaisen diskretoinnin käyttäjä voi suorittaa diskretisoinnin itse ennen aineiston syöttämistä ohjelmaan. [5, 8]

3.2 Mallin etsiminen

3.2.1 Mallien lukumäärä

Bayesin verkkojen avulla esitettävien riippuvuusmallien lukumäärä riippuu muuttujien määrästä, ja kasvaa nopeasti muuttujien määrän lisääntyessä [5]. Mallien lukumäärälle voidaan laskea suuntaa antava arvio kaavalla $2^{n(n-1)/2}$, kun muuttujia on n kappaletta. Tämä arvio on kuitenkin paljon todellisuutta pienempi, mutta kuvaa kuitenkin hyvin sitä, kuinka nopeasti mallien määrä kasvaa. Annettua kaavaa käyttämällä voidaan esimerkiksi todeta, että jo seitsemän muuttujan tapauksessa mahdollisia riippuvuusmalleja on yli kaksi miljoonaa. [9]

B-course-ohjelma huomioi vain ne mallit, jotka voidaan kuvata Bayesin verkoilla, ja tästä syystä moni riippuvuusmalli jää ohjelman toiminnan ulkopuolelle. Bayes-verkoilla kuvattavissa riippuvuusmalleissa muuttujien välisten suhteiden tulee olla kausaalisia. Toisinaan muuttujien välisen kausaalisen suhteen takana on latentti muuttuja, ja tätä tilannetta ei voida kuvata tarkasti bayes-verkoilla. Jos latenttia muuttujaa ei ole, muuttujien välinen suhde on kausaalinen. Latenttien muuttujien mahdollinen olemassaolo hankaloittaa riippuvuussuhteiden tulkintaa. [4, 5] Muuttujien välisiä suhteita esitellään tarkemmin luvussa 2.

3.2.2 Parhaan mallin etsiminen

B-course-ohjelmassa paras malli on se malli, jonka todennäköisyys on suurin. Ohjelma pystyy vertailemaan eri mallien todennäköisyyksiä bayesiläisen teorian avulla [4, 5]. Useimmiten muuttujia on paljon, jolloin vaihtoehtoisten mallien lukumäärä on todella suuri, ja kaikkien mallien läpikäynti kohtuullisessa ajassa on mahdotonta. Todennäköisimmän mallin löytäminen varmuudella on siis mahdollista vain silloin, kun muuttujien lukumäärä on hyvin pieni. B-course-ohjelman metodeilla löytyvä malli ei välttämättä aina ole todennäköisin malli, mutta kuitenkin sellainen malli, jonka todennäköisyys on lähes yhtä suuri kuin parhaan mallin todennäköisyys. Suurelle aineistolle B-course saattaa tarjota hieman toisistaan poikkeavia parhaita malleja eri ajokerroilla.

B-course suorittaa parhaan mallin etsinnän vertaamalla kahden mallin todennäköisyyksiä toisiinsa ja valitsemalla näistä sen, jonka todennäköisyys on suurempi. Seuraavaksi vertailuun valitaan malli, joka muistuttaa senhetkistä parasta mallia (eli edellisen parin voittajaa), ja näiden todennäköisyyksiä verrataan toisiinsa, ja jälleen parempi valitaan. Kun vertailtava malli valitaan tällä tavalla, löydetään todennäköisemmin malli, jonka todennäköisyys on suurempi kuin senhetkisen parhaan mallin todennäköisyys, kuin jos vertailuun valittaisiin malli satunnaisesti. [4]

Jos monilla malleilla on lähes sama todennäköisyys kuin parhaalla mallilla, tulisi myös nämä mallit huomioida, kun parhaan mallin pohjalta tehdään päätelmiä. Vain yhden mallin perusteella tehdyt päätelmät eivät ole yhtä päteviä kuin usean mallin perusteella tehdyt päätelmät. Jos taas paras malli on paljon todennäköisempi kuin muut mallit, voidaan päätelmät tehdä vain tämän yhden mallin perusteella. Tämän selvittämiseksi data tulee analysoida ohjelmalla useamman kerran, ja jos joka kerta tai lähes joka kerta tuloksena on sama malli, sen todennäköisyys on todennäköisesti selkeästi suurempi kuin muiden mallien. Jos tuloksena on joka kerta erilainen malli, päätelmien tekemisessä olisi järkevää hyödyntää useampaa kuin yhtä mallia.

3.2.3 Riippuvuusmallien todennäköisyydet

Bayesin säännön avulla voidaan laskea riippuvuusmallin todennäköisyys tietylle aineistolle. Mallin todennäköisyys on siis ehdollinen ja riippuu tutkittavasta aineistosta. Kun on olemassa data D , lasketaan todennäköisyys riippuvuusmallille M käyttäen Bayesin kaavaa (2.4)

$$P(M|D) = P(D|M) \frac{P(M)}{P(D)} \quad (3.1)$$

Bayesin säännössä $P(M|D)$ on datan D antama todennäköisyys mallille M ja $P(D|M)$ on mallin M antama todennäköisyys datalle D . $P(M)$ on mallin M toden-

näköisyys ilman dataa D ja $P(D)$ on datan D ”yleistodennäköisyys”. [3, 10]

Todennäköisyyttä $P(D)$ on hyvin työläs laskea, joten yksittäisen mallin todennäköisyyden laskeminen on lähes mahdotonta. Todennäköisimmän mallin etsiminen on kuitenkin mahdollista ilman, että yksittäisen mallin todennäköisyyttä lasketaan, sillä riittää, että pystytään vertailemaan kahden mallin todennäköisyyksiä keskenään. Laskettaessa mallien M_1 ja M_2 todennäköisyyksien $P(M_1|D)$ ja $P(M_2|D)$ suhde $P(D)$ supistuu pois:

$$\frac{P(M_1|D)}{P(M_2|D)} = \frac{P(D|M_1)P(M_1)/P(D)}{P(D|M_2)P(M_2)/P(D)} = \frac{P(D|M_1)P(M_1)}{P(D|M_2)P(M_2)} \quad (3.2)$$

Jos suhde $P(M_1|D)/P(M_2|D)$ suurempi kuin yksi, mallin M_1 todennäköisyys on suurempi kuin mallin M_2 todennäköisyys. Jos taas $0 < P(M_1|D)/P(M_2|D) < 1$, mallin M_2 todennäköisyys on suurempi kuin mallin M_1 todennäköisyys.

Suhteen $P(M_1|D)/P(M_2|D)$ laskemiseksi on tiedettävä, kuinka todennäköisyydet $P(M)$ ja $P(D|M)$ lasketaan. $P(M)$ on mallin M todennäköisyys ilman dataa, eli mallin M prioritodennäköisyys. Bayesiläisessä tilastotieteessä prioritodennäköisyys pyritään määrittämään tiedossa olevien taustatietojen avulla. B-course-ohjelma olettaa, että kaikki mallit ovat aluksi yhtä todennäköisiä, joten $P(M_1) = P(M_2)$ kaikille malleille M_1 ja M_2 . Tästä seuraa, että suhteesta $P(M_1|D)/P(M_2|D)$ supistuvat myös $P(M_1)$ ja $P(M_2)$ pois:

$$\frac{P(D|M_1)P(M_1)}{P(D|M_2)P(M_2)} = \frac{P(D|M_1)P(M_1)}{P(D|M_2)P(M_1)} = \frac{P(D|M_1)}{P(D|M_2)} \quad (3.3)$$

Mallien todennäköisyyksien suhteen selvittämiseksi tarvitsee lopulta laskea kummallekin mallille todennäköisyyden $P(D|M)$ arvo. Tämä todennäköisyys saadaan laskettua kaavalla

$$P(D|M) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\frac{N'}{q_i})}{\Gamma(\frac{N'}{q_i} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\frac{N'}{r_i q_i} + N_{ijk})}{\Gamma(\frac{N'}{r_i q_i})}, \quad (3.4)$$

missä

- n on muuttujien lukumäärä
- r_i on i :nnen muuttujan arvojen lukumäärä
- q_i on i :nnen muuttujan vanhempien mahdollisten konfiguraatioiden lukumäärä. i :nnen muuttujan vanhempia ovat ne muuttujat, joista on kaari mallin i :nteen muuttujaan. Vanhempien konfiguraatioiden lukumäärä tarkoittaa lukumäärää yhdistelmiä, joita vanhempien arvot voivat saada. q_i voidaan siis laskea kertomalla keskenään kaikkien vanhempien arvojen lukumäärä. Nämä

yhdistelmät on listattu ja niille on annettu indeksi j , joka saa arvot ensimmäisestä yhdistelmästä q_i :nteen arvojen yhdistelmään.

- Γ on *gammafunktio*

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt, \quad (3.5)$$

missä $x > 0$ [11]

- N_{ij} kertoo niiden rivien lukumäärän datassa, joilla on j :s konfiguraatio i :nnen muuttujan vanhemmista
- N_{ijk} kertoo niiden rivien lukumäärän datassa, joilla on k :s arvo i :nnelle muuttujalle, jolle on j :s konfiguraatio
- N' on ekvivalentti otoskoko. N' on parametri, joka määrittelee, kuinka helposti voidaan muuttaa uskomuksia riippuvuuksien kvantitatiivisesta luonteesta datan perusteella. Jos N' on pieni, niin datalla on suuri vaikutus uskomuksiin, ja jos N' on suuri, data ei kovin helposti vaikuta uskomuksiin. [5]

Parametrin N' valinta vaikuttaa mallien todennäköisyyksiin, joten sillä on vaikutusta myös tulokseen todennäköisimmän mallin etsinnässä. B-course-ohjelmassa parametrin N' valinnassa on vaikuttaneita Jeffreysin priorista. *Jeffreysin prior* on bayesiläisessä tilastotieteessä esiintyvä epäinformatiivinen prior. Epäinformatiivinen prior kuvaa tietämättömyyttä hypoteesien suhteellisista todennäköisyyksistä. Jeffreysin prior parametrille θ määritellään kaavalla

$$\pi(\theta) = \frac{|I(\theta)|^{1/2}}{\int |I(\eta)|^{1/2} d\eta}, \quad (3.6)$$

missä

$$[I(\theta)]_{ij} = - \int p(x|\theta) \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p(x|\theta) dx \quad (3.7)$$

on *Fisherin informaatiomatriisi*. [5, 8, 10, 12] Jeffreysin prior on invariantti muuttujien muutoksille, ja siksi se olisi sopiva prior bayesverkkojen todennäköisyyksien laskemisessa, mutta se on kuitenkin laskennallisesti vaikea käyttää. N' on pyritty valitsemaan siten, että sen jakauma muistuttaisi mahdollisimman paljon Jeffreysin priorin jakaumaa, mutta olisi kuitenkin laskennallisesti helpompi käyttää. Ekvivalentti otoskoko N' saadaan laskettua kaavalla

$$N' = \frac{\sum_{i=1}^n r_i}{2n}, \quad (3.8)$$

missä n on muuttujien lukumäärä ja r_i on i :nnen muuttujan arvojen lukumäärä. Tyhjälle bayesverkolle, eli verkolle, jossa ei ole kaaria, N' :lle saadaan kaavalla (3.8) sama arvo kuin Jeffreysin priorin kaavalla (3.6). [5, 8]

3.2.4 Esimerkki todennäköisyyden $P(D/M)$ laskemisesta

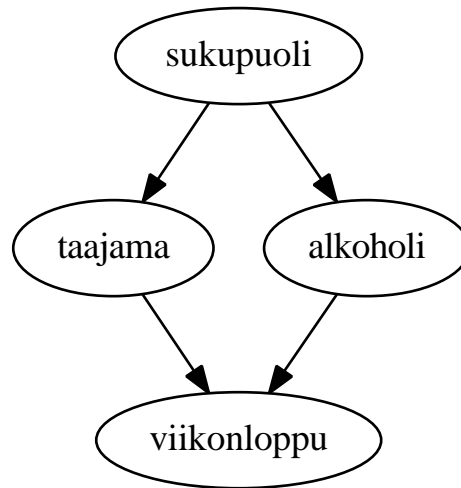
Lasketaan esimerkkiaineiston avulla kaavan 3.4 mukainen ehdollinen todennäköisyys aineistolle, kun riippuvuusmalli tiedetään. Käytetty aineisto (taulukko 3.1) sisältää tietoja liikenneonnettomuuksista, ja muuttujat ovat samoja tai lähes samoja kuin tässä työssä tutkittavan aineiston muuttujat, mutta aineiston arvot ovat keksittyjä. Muuttuja *sukupuoli* tarkoittaa kuljettajan sukupuolta, ja *alkoholi* sitä, onko onnettomuus ajettu alkoholin vaikutuksen alaisena. Muuttuja *taajama* kertoo, onko onnettomuus sattunut taajamassa, ja *viikonloppu*, onko onnettomuus sattunut viikonloppuna.

Taulukko 3.1: Esimerkin aineisto. Arvot ovat keksittyjä eivätkä vastaa työssä tutkittavaa aineistoa.

| <i>sukupuoli</i> | <i>taajama</i> | <i>viikonloppu</i> | <i>alkoholi</i> |
|------------------|----------------|--------------------|-----------------|
| nainen | ei | ei | ei |
| mies | kyllä | ei | ei |
| mies | ei | ei | ei |
| nainen | ei | ei | ei |
| mies | ei | kyllä | kyllä |
| mies | ei | kyllä | kyllä |
| nainen | ei | kyllä | ei |
| mies | kyllä | ei | kyllä |
| mies | ei | kyllä | ei |
| nainen | ei | ei | ei |
| mies | ei | ei | ei |
| mies | kyllä | ei | ei |

Aineiston muuttujien määrä on valittu pieneksi, ja muuttujat on lisäksi valittu siten, että jokaisella on vain kaksi mahdollista arvoa. Näitten valintojen tarkoituksena on välttyä liian pitkiltä ja monimutkaisilta laskuilta, ja pitää esimerkki mahdollisimman selkeänä.

B-course-ohjelmistolla aineistolle saadaan kuvan 3.1 mukainen riippuvuusmalli, jota käytetään todennäköisyyden $P(D|M)$ laskemiseen kaavalla (3.4).



Kuva 3.1: B-course-ohjelman antama paras riippuvuusmalli esimerkin aineistolle.

Olkoon $i = 1$ muuttujalle *sukupuoli*, $i = 2$ muuttujalle *taajama*, $i = 3$ muuttujalle *viikonloppu* ja $i = 4$ muuttujalle *alkoholi*.

- muuttujien lukumäärä $n = 4$
- muuttujien arvojen lukumäärät:
 - $r_1 = 2$
 - $r_2 = 2$
 - $r_3 = 2$
 - $r_4 = 2$
- muuttujien vanhempien konfiguraatioiden lukumäärät:
 - $q_1 = 1$, koska muuttujalla *sukupuoli* ei ole vanhempia, niin mahdollisia konfiguraatioita on yksi
 - $q_2 = r_1 = 2$
 - $q_3 = r_2 r_4 = 4$
 - $q_4 = r_1 = 2$
- ekvivalentti otoskoko $N' = \frac{\sum_{i=1}^n r_i}{2n} = \frac{2+2+2+2}{2 \cdot 4} = 1$

Todennäköisyyden $P(D|M)$ laskeminen yhtenä kokonaisuutena on pitkä ja monimutkainen lasku, joten jaetaan sen laskeminen pienempiin osakokonaisuuksiin. Jotta laskeminen voidaan suorittaa vaiheittain jaetaan kaava 3.4 osiin:

$$P(D|M) = \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\frac{N'}{q_i})}{\Gamma(\frac{N'}{q_i} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\frac{N'}{r_i q_i} + N_{ijk})}{\Gamma(\frac{N'}{r_i q_i})} = \prod_{i=1}^4 P_i, \quad (3.9)$$

missä

$$P_i = \prod_{j=1}^{q_i} \frac{\Gamma(\frac{N'}{q_i})}{\Gamma(\frac{N'}{q_i} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\frac{N'}{r_i q_i} + N_{ijk})}{\Gamma(\frac{N'}{r_i q_i})}. \quad (3.10)$$

Lasketaan seuraavaksi P_i :t erikseen.

1. Osatodennäköisyys P_1 :

Olkoon muuttujan *sukupuoli* konfiguraatiot $j = 1$ arvolle *mies* ja $j = 2$ arvolle *nainen*, sekä $k = 1$ muuttujan *sukupuoli* arvolle *mies* ja $k = 2$ muuttujan *sukupuoli* arvolle *nainen*.

Nyt, kun $i = 1$, $N_{11}=12$ (kaikki datarivit, koska muuttujalla *sukupuoli* ei ole vanhempia), $N_{111}=8$ (kaikki ne rivit, joissa sukupuoli on *mies*) $N_{112}=4$ (kaikki ne rivit, joissa sukupuoli on *nainen*), saadaan:

$$\begin{aligned} P_1 &= \prod_{j=1}^{q_1=1} \frac{\Gamma(\frac{N'}{q_1})}{\Gamma(\frac{N'}{q_1} + N_{1j})} \prod_{k=1}^{r_1=2} \frac{\Gamma(\frac{N'}{r_1 q_1} + N_{1jk})}{\Gamma(\frac{N'}{r_1 q_1})} \\ &= \frac{\Gamma(\frac{1}{1})}{\Gamma(\frac{1}{1} + N_{11})} \left[\frac{\Gamma(\frac{1}{1.2} + N_{111})}{\Gamma(\frac{1}{1.2})} \cdot \frac{\Gamma(\frac{1}{1.2} + N_{112})}{\Gamma(\frac{1}{1.2})} \right] \\ &= \frac{\Gamma(1)}{\Gamma(1 + 12)} \left[\frac{\Gamma(\frac{1}{2} + 8)}{\Gamma(\frac{1}{2})} \cdot \frac{\Gamma(\frac{1}{2} + 4)}{\Gamma(\frac{1}{2})} \right] = \frac{\Gamma(1)}{\Gamma(13)} \left[\frac{\Gamma(8\frac{1}{2})}{\Gamma(\frac{1}{2})} \cdot \frac{\Gamma(4\frac{1}{2})}{\Gamma(\frac{1}{2})} \right] \end{aligned}$$

2. Osatodennäköisyys P_2 :

Taulukko 3.2: Muuttujien *sukupuoli* ja *taajama* arvojen yhdistelmät, ja näiden yhdistelmien lukumäärät (N_{ijk} :t) aineistossa.

| N_{ijk} | <i>sukupuoli</i> | <i>taajama</i> | N_{ijk} :n arvo |
|-----------|------------------|----------------|-------------------|
| N_{211} | mies | kyllä | 3 |
| N_{212} | mies | ei | 5 |
| N_{221} | nainen | kyllä | 0 |
| N_{222} | nainen | ei | 4 |

Kun $i = 2$, $N_{21}=8$ (rivit, joissa sukupuoli on *mies*), $N_{22}=4$ (rivit, joissa sukupuoli on *nainen*) ja N_{ijk} :n arvot ovat taulukon 3.2 mukaiset:

$$\begin{aligned}
P_2 &= \prod_{j=1}^{q_2=2} \frac{\Gamma(\frac{N'}{q_2})}{\Gamma(\frac{N'}{q_2} + N_{2j})} \prod_{k=1}^{r_2=2} \frac{\Gamma(\frac{N'}{r_2 q_2} + N_{2jk})}{\Gamma(\frac{N'}{r_2 q_2})} \\
&= \left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2} + N_{21})} \cdot \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2} + N_{22})} \right] \\
&\quad \cdot \left[\frac{\Gamma(\frac{1}{2.2} + N_{211})}{\Gamma(\frac{1}{2.2})} \cdot \frac{\Gamma(\frac{1}{2.2} + N_{212})}{\Gamma(\frac{1}{2.2})} \cdot \frac{\Gamma(\frac{1}{2.2} + N_{221})}{\Gamma(\frac{1}{2.2})} \cdot \frac{\Gamma(\frac{1}{2.2} + N_{222})}{\Gamma(\frac{1}{2.2})} \right] \\
&= \left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2} + 8)} \cdot \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2} + 4)} \right] \cdot \left[\frac{\Gamma(\frac{1}{4} + 3)}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4} + 5)}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4} + 0)}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4} + 4)}{\Gamma(\frac{1}{4})} \right] \\
&= \left[\frac{\Gamma(\frac{1}{2})}{\Gamma(8\frac{1}{2})} \cdot \frac{\Gamma(\frac{1}{2})}{\Gamma(4\frac{1}{2})} \right] \cdot \left[\frac{\Gamma(3\frac{1}{4})}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(5\frac{1}{4})}{\Gamma(\frac{1}{4})} \cdot 1 \cdot \frac{\Gamma(4\frac{1}{4})}{\Gamma(\frac{1}{4})} \right]
\end{aligned}$$

3. Osatodennäköisyys P_3 :

Taulukko 3.3: Konfiguraatiot muuttujan *viikonloppu* vanhemmille.

| j | <i>taajama</i> | <i>alkoholi</i> | N_{ij} | N_{ij} :n arvo |
|-----|----------------|-----------------|----------|------------------|
| 1 | kyllä | kyllä | N_{31} | 1 |
| 2 | kyllä | ei | N_{32} | 2 |
| 3 | ei | kyllä | N_{33} | 2 |
| 4 | ei | ei | N_{34} | 7 |

Taulukko 3.4: Muuttujien *taajama*, *alkoholi* ja *viikonloppu* arvojen yhdistelmät, ja näiden yhdistelmien lukumäärät (N_{ijk} :t) aineistossa.

| N_{ijk} | <i>taajama</i> | <i>alkoholi</i> | <i>viikonloppu</i> | N_{ijk} :n arvo |
|-----------|----------------|-----------------|--------------------|-------------------|
| N_{311} | kyllä | kyllä | kyllä | 0 |
| N_{312} | kyllä | kyllä | ei | 1 |
| N_{321} | kyllä | ei | kyllä | 0 |
| N_{322} | kyllä | ei | ei | 2 |
| N_{331} | ei | kyllä | kyllä | 2 |
| N_{332} | ei | kyllä | ei | 0 |
| N_{341} | ei | ei | kyllä | 2 |
| N_{342} | ei | ei | ei | 5 |

Kun $i = 3$, N_{ij} :n arvot ovat taulukon 3.3 ja N_{ijk} :n arvot ovat taulukon 3.4 mukaiset:

$$\begin{aligned}
P_3 &= \prod_{j=1}^{q_3=4} \frac{\Gamma(\frac{N'}{q_3})}{\Gamma(\frac{N'}{q_3} + N_{3j})} \prod_{k=1}^{r_3=2} \frac{\Gamma(\frac{N'}{r_3 q_3} + N_{3jk})}{\Gamma(\frac{N'}{r_3 q_3})} \\
&= \left[\frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{4} + N_{31})} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{4} + N_{32})} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{4} + N_{33})} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{4} + N_{34})} \right] \\
&\quad \cdot \left[\frac{\Gamma(\frac{1}{2.4} + N_{311})}{\Gamma(\frac{1}{2.4})} \cdot \frac{\Gamma(\frac{1}{2.4} + N_{312})}{\Gamma(\frac{1}{2.4})} \cdot \frac{\Gamma(\frac{1}{2.4} + N_{321})}{\Gamma(\frac{1}{2.4})} \cdot \frac{\Gamma(\frac{1}{2.4} + N_{322})}{\Gamma(\frac{1}{2.4})} \right. \\
&\quad \cdot \frac{\Gamma(\frac{1}{2.4} + N_{331})}{\Gamma(\frac{1}{2.4})} \cdot \frac{\Gamma(\frac{1}{2.4} + N_{332})}{\Gamma(\frac{1}{2.4})} \cdot \frac{\Gamma(\frac{1}{2.4} + N_{341})}{\Gamma(\frac{1}{2.4})} \cdot \left. \frac{\Gamma(\frac{1}{2.4} + N_{342})}{\Gamma(\frac{1}{2.4})} \right] \\
&= \left[\frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{4} + 1)} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{4} + 2)} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{4} + 2)} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(\frac{1}{4} + 7)} \right] \cdot \left[\frac{\Gamma(\frac{1}{8} + 0)}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(\frac{1}{8} + 1)}{\Gamma(\frac{1}{8})} \right. \\
&\quad \cdot \frac{\Gamma(\frac{1}{8} + 0)}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(\frac{1}{8} + 2)}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(\frac{1}{8} + 2)}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(\frac{1}{8} + 0)}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(\frac{1}{8} + 2)}{\Gamma(\frac{1}{8})} \cdot \left. \frac{\Gamma(\frac{1}{8} + 5)}{\Gamma(\frac{1}{8})} \right] \\
&= \left[\frac{\Gamma(\frac{1}{4})}{\Gamma(1\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(2\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(2\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(7\frac{1}{4})} \right] \\
&\quad \cdot \left[1 \cdot \frac{\Gamma(1\frac{1}{8})}{\Gamma(\frac{1}{8})} \cdot 1 \cdot \frac{\Gamma(2\frac{1}{8})}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(2\frac{1}{8})}{\Gamma(\frac{1}{8})} \cdot 1 \cdot \frac{\Gamma(2\frac{1}{8})}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(5\frac{1}{8})}{\Gamma(\frac{1}{8})} \right]
\end{aligned}$$

4. Osatodennäköisyys P_4 :

Taulukko 3.5: Muuttujien *sukupuoli* ja *alkoholi* arvojen yhdistelmät, ja näiden yhdistelmien lukumäärät (N_{ijk} :t) aineistossa.

| N_{ijk} | <i>sukupuoli</i> | <i>alkoholi</i> | N_{ijk} :n arvo |
|-----------|------------------|-----------------|-------------------|
| N_{411} | mies | kyllä | 3 |
| N_{412} | mies | ei | 5 |
| N_{421} | nainen | kyllä | 0 |
| N_{422} | nainen | ei | 4 |

Kun $i = 4$, $N_{41}=8$ (rivit, joissa sukupuoli on *mies*), $N_{42}=4$ (rivit, joissa sukupuoli on *nainen*) ja N_{ijk} :n arvot ovat taulukon 3.5 mukaiset:

$$\begin{aligned}
P_4 &= \prod_{j=1}^{q_4=2} \frac{\Gamma(\frac{N'}{q_4})}{\Gamma(\frac{N'}{q_4} + N_{4j})} \prod_{k=1}^{r_4=2} \frac{\Gamma(\frac{N'}{r_4 q_4} + N_{4jk})}{\Gamma(\frac{N'}{r_4 q_4})} \\
&= \left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2} + N_{41})} \cdot \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2} + N_{42})} \right] \\
&\quad \cdot \left[\frac{\Gamma(\frac{1}{2.2} + N_{411})}{\Gamma(\frac{1}{2.2})} \cdot \frac{\Gamma(\frac{1}{2.2} + N_{412})}{\Gamma(\frac{1}{2.2})} \cdot \frac{\Gamma(\frac{1}{2.2} + N_{421})}{\Gamma(\frac{1}{2.2})} \cdot \frac{\Gamma(\frac{1}{2.2} + N_{422})}{\Gamma(\frac{1}{2.2})} \right] \\
&= \left[\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2} + 8)} \cdot \frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{1}{2} + 4)} \right] \cdot \left[\frac{\Gamma(\frac{1}{4} + 3)}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4} + 5)}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4} + 0)}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4} + 4)}{\Gamma(\frac{1}{4})} \right] \\
&= \left[\frac{\Gamma(\frac{1}{2})}{\Gamma(8\frac{1}{2})} \cdot \frac{\Gamma(\frac{1}{2})}{\Gamma(4\frac{1}{2})} \right] \cdot \left[\frac{\Gamma(3\frac{1}{4})}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(5\frac{1}{4})}{\Gamma(\frac{1}{4})} \cdot 1 \cdot \frac{\Gamma(4\frac{1}{4})}{\Gamma(\frac{1}{4})} \right]
\end{aligned}$$

Lopuksi kertomalla P_i :t keskenään (kaava (3.9)), saadaan ehdollinen todennäköisyys aineistolle, kun malli tiedetään (gammafunktion arvot lasketaan Matlab-ohjelmistolla):

$$\begin{aligned}
P(D|M) &= \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{\Gamma(\frac{N'}{q_i})}{\Gamma(\frac{N'}{q_i} + N_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(\frac{N'}{r_i q_i} + N_{ijk})}{\Gamma(\frac{N'}{r_i q_i})} = \prod_{i=1}^4 P_i \\
&= \left[\frac{\Gamma(1)}{\Gamma(13)} \left[\frac{\Gamma(8\frac{1}{2})}{\Gamma(\frac{1}{2})} \cdot \frac{\Gamma(4\frac{1}{2})}{\Gamma(\frac{1}{2})} \right] \right] \cdot \left[\left[\frac{\Gamma(\frac{1}{2})}{\Gamma(8\frac{1}{2})} \cdot \frac{\Gamma(\frac{1}{2})}{\Gamma(4\frac{1}{2})} \right] \cdot \left[\frac{\Gamma(3\frac{1}{4})}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(5\frac{1}{4})}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(4\frac{1}{4})}{\Gamma(\frac{1}{4})} \right] \right] \\
&\quad \cdot \left[\left[\frac{\Gamma(\frac{1}{4})}{\Gamma(1\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(2\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(2\frac{1}{4})} \cdot \frac{\Gamma(\frac{1}{4})}{\Gamma(7\frac{1}{4})} \right] \cdot \left[\frac{\Gamma(1\frac{1}{8})}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(2\frac{1}{8})}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(2\frac{1}{8})}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(2\frac{1}{8})}{\Gamma(\frac{1}{8})} \cdot \frac{\Gamma(5\frac{1}{8})}{\Gamma(\frac{1}{8})} \right] \right] \\
&\quad \cdot \left[\left[\frac{\Gamma(\frac{1}{2})}{\Gamma(8\frac{1}{2})} \cdot \frac{\Gamma(\frac{1}{2})}{\Gamma(4\frac{1}{2})} \right] \cdot \left[\frac{\Gamma(3\frac{1}{4})}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(5\frac{1}{4})}{\Gamma(\frac{1}{4})} \cdot \frac{\Gamma(4\frac{1}{4})}{\Gamma(\frac{1}{4})} \right] \right] \\
&= \frac{(\Gamma(1)) \cdot (\Gamma(3\frac{1}{4}))^2 \cdot (\Gamma(5\frac{1}{4}))^2 \cdot (\Gamma(4\frac{1}{4}))^2 \cdot \Gamma(1\frac{1}{8}) \cdot (\Gamma(2\frac{1}{8}))^3 \cdot \Gamma(5\frac{1}{8}) \cdot (\Gamma(\frac{1}{2}))^2}{\Gamma(13) \cdot \Gamma(1\frac{1}{4}) \cdot (\Gamma(2\frac{1}{4}))^2 \cdot (\Gamma(\frac{1}{8}))^5 \cdot \Gamma(8\frac{1}{2}) \cdot \Gamma(4\frac{1}{2}) \cdot (\Gamma(\frac{1}{4}))^2 \cdot \Gamma(7\frac{1}{4})} \\
&= 1.6838 \cdot 10^{-15}
\end{aligned}$$

Saatu tulos ei ole mallin todennäköisyys vaan aineiston ehdollinen todennäköisyys kun malli tunnetaan. Paras malli etsitään vertailemalla aineiston ehdollista todennäköisyyttä eri malleille kaavan (3.3) mukaisesti. Mitä suurempi $P(D|M)$ on, sitä paremmin kyseinen malli kuvaa aineistoa.

4. TIEDOT LIIKENNEONNETTOMUUKSISTA PIRKANMAALLA VUOSINA 2004–2008

Pirkanmaan liikenneonnettomuudet vuosien 2004 – 2008 aikana on osa aineistoa, jossa on tiedot koko Suomen alueella vuosien 2004 – 2008 aikana sattuneista liikenneonnettomuuksista. Tätä työtä varten on rajattu vain osa koko aineistosta, sillä työssä käytettävään B-course-ohjelmistoon on mahdollista syöttää korkeintaan vain yhden megatavun kokoinen tiedosto. Alkuperäisessä koko Suomen kattavassa liikenneonnettomuusaineistossa on suuremman liikenneonnettomuusmäärän lisäksi myös muuttujia enemmän kuin tässä työssä käytetyssä rajatussa aineistossa.

Työssä käytetty Pirkanmaan liikenneonnettomuudet -aineisto sisältää tiedot 8676 liikenneonnettomuudesta. Muuttujia aineistossa on 44. Jokaisesta aineiston onnettomuudesta ei ole täydellistä dataa eli joidenkin muuttujien osalta puuttuu tietoja. Tämä johtuu siitä, että osa muuttujien tiedoista ei ole ollut relevanttia kyseisessä onnettomuudessa tai tietoa ei ole muuten ollut mahdollista saada. Yleisin syy tähän on se, että suuri osa muuttujista sisältää tietoja risteysalueista, ja jos liikenneonnettomuus ei ole tapahtunut tällaisella alueella, kyseisiä tietoja ei siten ole olemassa.

Aineiston muuttujat kuvaavat onnettomuuden aiheuttanutta kuljettajaa, sääolosuhteita, onnettomuuden ajankohtaa ja tapahtumapaikkaa sekä onnettomuuden seurauksia. Muuttujat on esitelty ryhmittäin, ja muuttujan esittelyn yhteydessä on maininta, jos merkittävää osuutta muuttujan arvoista ei ole olemassa, ja myös kerrottu mahdollinen syy siihen, miksi kyseiset arvot puuttuvat. Suurin osa muuttujista on luokkamuuttujia. Aineistossa näiden muuttujien arvot on korvattu kokonaisluvuilla, joita vastaavat luokat on esitelty muuttujan esittelyn yhteydessä.

4.1 Tietoja kuljettajasta, keliolosuhteista ja onnettomuusajankohdasta

Onnettomuuden aiheuttaneeseen kuljettajaan liittyvät muuttujat ovat *ika*, *sukup*, *kulj-alkoh* ja *promil*. *Ika* tarkoittaa kuljettajan ikää vuosina. Muuttuja saa kuljettajan iän mukaisen kokonaislukuarvon, ja 98-vuotiaat ja sitä vanhemmat saa arvoksi 98. *Sukup* vastaan kuljettajan sukupuolta ja sen arvot ovat *M* (mies) ja *N* (nainen). *Kulj-alkoh* muuttuja saa arvon *On* tai *Ei*, ja kertoo, onko kuljettaja ollut onnet-

tomuushetkellä alkoholin vaikutuksen alaisena. Muuttuja *promil* sen sijaan kertoo kuljettajan puhaltaman promillelukeman desimaalilukuna, ja jos alkoholinkäytöstä epäillään, mutta puhallutustuloksia ei ole olemassa muuttujan arvona on 7.77. Noin joka toisen onnettomuuden osalta puuttuu tieto kuljettajan puhaltamasta promillelukemasta.

Sääolosuhteita kuvaavat muuttujat *saa*, *lampotila*, *pinta* ja *valoisuus*. *Lampotilan* arvot ovat kokonaislukuja ja vastaavat ilman lämpötilaa tapahtumapaikassa tapahtumahetkellä. Muuttujalla *pinta* tarkoitetaan tien pintaa onnettomuuspaikalla. *Saa*, *pinta* ja *valoisuus* kuvaavat vallitsevia olosuhteita, ja niiden saamien arvojen selitykset ovat taulukossa 4.1.

Keliolosuhteista onnettomuuspaikalla kertovat muuttujat *kp-luokka*, *paal-lk* ja *paallyste*. Muuttuja *kp-luokka* kertoo tien hoitoluokan (taulukko 4.2). Tien hoitoluokalla on merkitystä tien kuntoon lähinnä talven aikana. Muuttujat *paal-lk* ja *paallys-te* kertovat tien päällystemateriaalista taulukon 4.3 mukaisesti.

Taulukko 4.1: Muuttujien *saa*, *pinta* ja *valoisuus* arvojen selitykset.

| Arvo | Sää | Pinta | Valoisuus |
|------|------------|-----------------|----------------------|
| 1 | kirkas | paljas, kuiva | päivänvalo |
| 2 | pilvipouta | paljas, märkä | hämärä |
| 3 | sumu | urissa vettä | pimeä (valaisematon) |
| 4 | vesisade | luminen | tie valaistu |
| 5 | lumisade | sohjoinen | |
| 6 | räntäsade | jäinen | |
| 7 | raesade | ajourat paljaat | |

Taulukko 4.2: Muuttujan *kp-luokka* arvoja vastaavat tien hoitoluokat.

| Arvo | Hoitoluokka |
|------|---|
| 1 | 1s normaalisti aina paljaina |
| 2 | 1 tingitään öisin |
| 3 | 1b osan talvea lumipintaisena |
| 4 | T1b 1b-luokka taajaman alueella |
| 5 | II pääosin lumipintainen |
| 6 | III lumipintainen, hiekoitus vain pahimmissa olosuhteissa |
| 7 | K1 Hyvin hoidettu kevyen liikenteen väylä |
| 8 | K2 Vähämerkityksinen kevyen liikenteen väylä |

Onnettomuusajankohtaa kuvaavat muuttujat *tunti*, *vkpv* ja *kk*. *Tunti* vastaa kellonaikaa, ja saa kokonaislukuarvoja välillä 0 – 23, *vkpv* tarkoittaa viikonpäivää, ja sen arvot ovat *ma*, *ti*, *ke*, *to*, *pe*, *la* ja *su*. Muuttuja *kk* vastaa kuukautta ja saa arvot 1 – 12.

Taulukko 4.3: Muuttujien *paallyste* ja *paal-lk* arvojen selitykset.

| Arvo | Päällyste | Arvo | Päällysteluokka |
|------|-----------------------|------|-------------------------|
| 1 | kestopäällyste | 1 | betoni |
| 2 | öljysora tai vastaava | 2 | kivi |
| 3 | sora | 10 | kovat asfalttibetonit |
| 4 | kivi | 20 | pehmeät asfalttibetonit |
| 5 | betoni | 30 | soratien pinta |
| 6 | muu | 40 | sorakulutuskerros |

4.2 Onnettomuuden tyyppi ja luokittelu

Onnettomuudessa kuolleiden määrä ilmoitetaan datassa muuttujalla *kuolleet* ja loukkaantuneiden määrä muuttujalla *loukkaant.* Onnettomuuteen joutuneiden osapuolten lukumäärän ilmaisee muuttuja *osallkm.* Muuttuja *rask-osall* kertoo, onko onnettomuuden osapuolena ollut raskas kulkuneuvo (*On*) vai ei (*Ei*). Henkilö- ja ajoneuvovahinkojen lisäksi mahdollisista muista omaisuusvahingoista kertoo muuttuja *muu vah.*, jonka arvojen selitykset on esitetty taulukossa 4.4.

Taulukko 4.4: Muiden omaisuusvahinkojen selitykset muuttujalle *muu-vah.*

| Arvo | Muu omaisuusvahinko | Arvo | Muu omaisuusvahinko |
|------|-----------------------------|------|---------------------|
| 0 | ei muuta omaisuusvahinkoa | 4 | liikennekoroke |
| 1 | liikennemerkki | 5 | muita vaurioita |
| 2 | valaisinpylväs tai vastaava | 6 | siltapilari |
| 3 | reuna- tai suojakaide | 7 | liikennevalopylväs |

Taulukko 4.5: Onnettomuusluokat muuttujalle *onluokka*.

| Arvo | Onnettomuusluokka | Arvo | Onnettomuusluokka |
|------|-----------------------|------|-------------------------|
| 1 | yksittäisonnettomuus | 8 | polkupyöräonnettomuus |
| 2 | kääntymisonnettomuus | 9 | jalankulkijaonnettomuus |
| 3 | ohitusonnettomuus | 10 | hirvionnettomuus |
| 4 | risteämisonnettomuus | 11 | peuraonnettomuus |
| 5 | kohtaamisonnettomuus | 12 | muu eläinonnettomuus |
| 6 | peräänajo-onnettomuus | 13 | muu onnettomuus |
| 7 | mopedionnettomuus | | |

Onnettomuuden tyyppiä ja luokkaa kuvaavat muuttujat *ontyyppi* ja *onluokka*. Muuttuja *ontyyppi* kuvaa onnettomuustyyppejä, joita on 76 kappaletta. *ontyyppi*-muuttujan arvoa vastaavat numerokoodit on lueteltu liitteessä A. *onluokka*-muuttuja vastaa onnettomuusluokkaa, joka määräytyy osallisten lukumäärän, onnettomuuslajin ja onnettomuustyyppin perusteella. Onnettomuusluokat on esitelty taulukossa 4.5.

4.3 Tietoja onnettomuuspaikasta

Muuttuja *taaj-merk* ilmaisee, onko liikenneonnettomuus tapahtunut taajamassa (K) vai taajama-alueen ulkopuolella (E). *asutus*-muuttuja taas kertoo taulukon 4.6 mukaisesti, kuinka tiheään asutulla alueella onnettomuuspaikka sijaitsee. Onnettomuuspaikan tien vilkkautta kuvaavat muuttujat *kvl* ja *raskas-kvl*, jotka kertovat keskimääräisen vuorokausiliikenteen ja raskaiden kulkuneuvojen keskimääräisen vuorokausiliikenteen. Muuttujat kertovat siis vuorokauden aikana onnettomuuspaikan ohittaneiden kulkuneuvojen lukumäärän.

Taulukko 4.6: Asutustiheyden luokittelu muuttujan *asutus* arvoille.

| Arvo | Asutustiheys |
|------|---|
| 1 | taajaman liikennemerkki |
| 2 | tilastollinen taajama |
| 3 | taajama, yli 60 asukasta/km ² |
| 4 | taajama, 30 – 60 asukasta/km ² |
| 5 | taajama 15 – 30 asukasta/km ² |
| 6 | harvaan asuttu alue, alle 15 asukasta/km ² |

Onnettomuusnopeuksia kuvaavat muuttujat *nop-suuntaan* ja *nop-vastaan*. Ensimmäinen muuttuja kertoo onnettomuuden aiheuttaneen kulkuneuvon nopeuden ja jälkimmäinen mahdollisen vastaan tulleen ja onnettomuuteen joutuneen kulkuneuvon nopeuden. Nopeudet ovat km/h ja pyöristetty kymmenlukuiksi. Noin puolesta onnettomuuksista puuttuu tieto vastaan tulleen nopeudesta, koska kyseisessä onnettomuudessa ei ollut vastaan tulevaa liikennettä osallisena.

Nopeusrajoituksista kertovat muuttujat *nop-raj* (nopeusrajoitus) ja *nop-laji* (nopeusrajoituslaji). Nopeusrajoitukset ovat todellisen tilanteen mukaisia nopeusrajoituksia, esimerkiksi talvella talvinopeusrajoituksen mukainen ja tietyömaalla työmaan aikainen nopeusrajoitus. Nopeusrajoituslaji kertoo vallitsevan nopeusrajoituksen voimassaoloalueen. Nopeusrajoituslajit on esitelty taulukossa 4.7.

Taulukko 4.7: Nopeusrajoituslajit, jotka vastaavat muuttujan *nop-laji* arvoja.

| Arvo | Nopeusrajoituslaji |
|------|-----------------------------|
| 1 | tiekohtainen rajoitus |
| 2 | pistekohtainen rajoitus |
| 3 | nopeusrajoituksen porrastus |
| 4 | paikallinen rajoitus |
| 5 | nopeusrajoitusalue |
| 6 | yleisrajoitus (taajamassa) |

Onnettomuuspaikan tien ominaisuuksia kuvaavat muuttujat *mo/mol*, *toim-luokka*, *tien-lev* ja *ajor-lkm*. Muuttuja *mo/mol* kertoo, onko onnettomuuspaikan tie moottoritie tai moottoritien ramppi (arvo 1), moottoriliikennetie tai moottoriliikennetien

ramppi (arvo 2) vai muu vain tietynlaista liikennettä varten tarkoitettu ajotie (arvo 3). Tämä tieto on noin kymmenellä prosentilla muuttujista. Tieluokkaa kuvaa muuttuja *toim-luokka*, jonka arvoina ovat 1 (valtatie), 2 (kantatie), 3 (seututie) ja 4 (yhdystie). Ajouradan leveyden kertoo *tien-lev*, jonka arvo on annettu desimetreinä ja *ajor-lkm* ilmaisee ajoratojen lukumäärän kyseisellä tiellä.

Muuttujat *nak-150*, *nak-300* ja *nak-460* kertovat näkemäprosentit 150, 300 ja 460 metrin etäisyyksillä. Näkemäprosentti 150 metrille määritetään laskemalla niiden tieosuuksien pituuksien summa, joilla näkemäpituus ylittää 150 metriä, ja jakamalla saatu tulos koko tieosan pituudella. Vastaavasti lasketaan 300 ja 460 metriä ylittävien näkemien prosenttiosuudet. [13] Muuttujien arvot ovat siis tällä tavalla laskettuja prosenttilukemia. Noin kolmasosasta onnettomuuksista puuttuu tiedot näkemäprosentista.

Taulukko 4.8: Muuttujan *onn-paikka* arvojen selitykset.

| Arvo | Onnettomuuspaikka | Arvo | Onnettomuuspaikka |
|------|-------------------------|------|-------------------------|
| 1 | ajorata | 6 | silta |
| 2 | suojatie | 7 | lautta tai lossipaikka |
| 3 | kevyen liikenteen väylä | 8 | eritasoliittymän ramppi |
| 4 | pysäköintialue tai piha | 9 | linja-autopysäkki |
| 5 | muu | | |

Onnettomuuden tapahtumapaikkaa kuvaavia muuttujia ovat *junction-type*, *solmun-tyyppi*, *onn-paikka*, *rist-var*, *tasol-tyyppi*, *liik-valot muu-liit* ja *suuntalkm*. Onnettomuuden tapahtumapaikkaa kuvaavan muuttujan *onn-paikka* arvot on esitetty taulukossa 4.8. Muuttujat *junction-type* ja *solmun-tyyppi* kertovat, onko onnettomuuspaikkana tieverkon solmukohta, sekä millaisesta solmusta on kysymys. Näiden muuttujien arvot on esitetty taulukossa 4.9, ja ne eroavat toisistaan vain siten, että muuttujan *junction-type* saadessa arvon 10 (ei solmua), *solmun-tyyppi* ei saa ollenkaan arvoa. Aineistossa suurella osalla onnettomuuksista ei ole arvoa muuttujalle *solmun-tyyppi*, sillä muuttujan *junction-type* arvojen perusteella neljä viidesosaa aineiston liikenneonnettomuuksista ei ole tapahtunut tieverkon solmukohdissa.

Jos onnettomuuspaikkana on liittymä, tarkempia tietoja siitä kertovat muuttujat *tasol-tyyppi*, *muu-liit*, *suuntalkm* ja *rist-var*. Muuttuja *tasol-tyyppi* kuvaa liittymän luokkaa ja muuttuja *muu-liit* kertoo muun mahdollisen liittymän tyyppin. Näiden kahden muuttujan arvot on esitetty taulukossa 4.10. Muuttuja *suuntalkm* kertoo, kuinka moneen suuntaan onnettomuuspaikalta kulkee ajorata. Muuttujat *tasol-tyyppi*, *muu-liit* ja *suuntalkm* saavat arvoja vain niiden onnettomuuksien osalta, jotka ovat tapahtuneet tieverkon solmukohdissa (alle viidesosa onnettomuuksista).

Taulukko 4.9: Muuttujien *junction-type* ja *solmun-tyyp* arvojen selitykset.

| Arvo | Junction type / Solmun tyyppi |
|------|---|
| 0 | yksityistie- / katuliittymä (erikoistapaus) |
| 1 | normaali tasoliittymä |
| 2 | pieni kiertoliittymä (keskisaarekkeen halkaisija alle 13 m tai josta ei vielä ole muodostettu liittymätietä) |
| 3 | suuri kiertoliittymä (kiertoliittymästä on muodostettu liittymätie ja mahdollisesti osaliittymiä) |
| 4 | y-liittymä (liittyvässä tiessä on haara > 50 m tai päätießä on vasemmalle kääntyviä varten oikealla odotustila) |
| 5 | eritasoliittymä |
| 6 | eritasoristeys (ei liittymismahdollisuutta) |
| 7 | yleinen tie / katu (kunnan ja yleisen tien raja) |
| 8 | muu jakopiste |
| 9 | tien kulkua selventävä apupiste |
| 10 | ei solmua (<i>solmun tyyppi ei saa tätä arvoa</i>) |

Taulukko 4.10: Muuttujien *muu-liit* ja *tasol-tyyppi* arvojen selitykset. *muu-liit* (muu liittymä) kuvaa liittymien erikoistapauksia ja *tasol-tyyppi* (liittymän luokka) kuvaa tavallisten tasoliittymien luokkia.

| Arvo | Muu liittymä | Liittymän luokka |
|------|-------------------------|---------------------------------|
| 0 | ei muuta liittymää | |
| 1 | eritasoliittymän ramppi | normaali tasoliittymä |
| 2 | levähdysalueen liittymä | |
| 3 | katu tai rakennuskaavat | kiertoliittymä |
| 4 | yksityistie tai -alue | osaliittymä ns. y-haaraliittymä |
| 5 | huoltoaseman liittymä | ramppliittymä |
| 6 | kevyen liikenteen väylä | |

Taulukko 4.11: Muuttujan *rist-var* arvojen selitykset, jotka kuvaavat onnettomuuspaikan liikenteenohjausta.

| Arvo | Ajojärjestys | Arvo | Ajojärjestys |
|------|-----------------------------------|------|----------------------------|
| 0 | linjaonnettomuus (ei liittymässä) | 5 | muu risteys |
| 1 | tasa-arvo | 6 | kärkikolmio, liikennevalot |
| 2 | kärkikolmio | 7 | STOP-merkki, liikennevalot |
| 3 | STOP-merkki | 8 | kiertoliittymä |
| 4 | liikennevalot | | |

Liittymän ajojärjestykseen vaikuttavista tekijöistä kertoo muuttuja *rist-var*, ja sen arvot on lueteltu taulukossa 4.11. Jos liittymässä on liikennevalot, niiden toimintaa kuvaa muuttuja *liik-valot*, jonka arvot ovat 1 (toiminnassa), 2 (vilkulla), 3 (ei toiminnassa) ja 4 (epäkunnossa). Vain niiden onnettomuuksien kohdalla on arvo *liik-valot*-muuttujalle, jotka ovat tapahtuneet liikennevaloristeyksessä (noin viisi prosenttia onnettomuuksista).

5. AINEISTON JAKAMINEN PIENEMPIIN KOKONAISUUKSIIN

44 muuttujan aineiston riippuvuusmalli on liian suuri tarkasteltavaksi kokonaisuena. Aineisto pitää siis jakaa osiin, jotta sitä voidaan analysoida paremmin. Aineistosta poistettiin joitakin muuttujia, joiden sisältämä tieto ei ole erityisen relevanttia liikenneonnettomuuksien kannalta tai valtaosasta onnettomuuksia muuttujalta puuttui arvo. Tässä luvussa on esitelty prosessi, jolla aineistoa on rajattu ja miten pienemmät muuttujakokonaisuudet on muodostettu. Lisäksi tässä luvussa esitellään näitä muuttujakokonaisuuksia vastaavat kausaaliset riippuvuusmallit.

5.1 Muuttujakokonaisuuksien rajaamisprosessi

5.1.1 Muuttujien karsiminen

Ennen varsinaista rajaamisprosessia aineistosta jätettiin muuttujat *solmun-tyyp* ja *tasol-tyyppi* pois, sillä niiden saadessa arvoja kummankin muuttujan arvot ovat samat kuin *junction-type*-muuttujan arvot (taulukot 4.9 ja 4.10), joten kyseiset muuttujat eivät tuo lisätietoa onnettomuudesta muuttujaan *junction-type* verrattuna. Myös muuttuja *paal-lk* jätettiin pois, koska sen sisältämät tiedot eivät merkittävästi eroa muuttujan *paallyste* tiedoista.

Lisäksi myös muuttuja *promil* jätettiin aineistosta pois, koska *promil*-muuttujalla on arvo vain noin puolella onnettomuuksista ja *kulj-alkoh*-muuttujalla on arvo jokaisen onnettomuuden kohdalla. Koska B-course-ohjelmisto muuttaa *promil*-muuttujan luokitteluasteikolliseksi, joten karkeasti ajateltuna *promil* vastaa muuttujaa *kulj-alkoh*, joka on luokitteluasteikollinen.

Työn tarkoituksena on ennemminkin testata B-course-ohjelmistoa eikä löytää kaikkia mahdollisia riippuvuussuhteita aineiston muuttujien välillä, joten muuttujat *liik-valot*, *mo/mol*, *muu-liit* ja *suuntalkm* jätettiin myös pois tutkittavasta aineistosta. Näiltä muuttujilta puuttui arvot yli 85 prosentilta onnettomuuksista. Tyhjät arvot olisi ollut mahdollista korvata ylimääräisellä arvolla tai antaa B-course-ohjelmiston käsitellä tyhjiä arvoja sisältävät rivit luvussa 3.1.1 esitellyn periaatteen mukaisesti. Ylimääräinen arvo olisi voinut olla esimerkiksi ”ei liikennevaloja” tai ”ei moottori(liikenne)tiellä”. Aineistoa tarkasteltaessa huomattiin, että aineistosta

selvästi puuttuu muutenkin muuttujilta arvoja, joten osa puuttuvista arvoista saattaisi olla muotoa ”puuttuva arvo”. Koska tyhjiä arvoja oli kuitenkin paljon, eikä ollut varmuutta, ovatko muuttujien tyhjät arvot jätetty pois, koska sopivaa vaihtoehtoa ei ole, vai puuttuuko muuttujilta arvoja jostakin muusta syystä, niin edellä mainitut muuttujat poistettiin tutkittavasta aineistosta.

5.1.2 Muuttujien arvojen muokkaaminen

B-course-ohjelmisto ryhmittelee numeerisia arvoja saavat muuttujan arvot kolmeen ryhmään, jos arvoja on vähintään seitsemän. Tämän takia joidenkin muuttujien arvoja muokattiin, jotta tulosten tulkinta olisi mielekkäämpää. Yli 90 prosentissa liikenneonnettomuuksista osallisia on yksi tai kaksi, noin joka viidennessä onnettomuudessa loukkaantuu ihmisiä, ja kuolonuhreja tulee vain yhdessä prosentissa onnettomuuksia. Muuttujien *osallkm*, *loukkaant* ja *kuolleet* arvoja ryhmiteltiin siten, että pienimmät arvot ovat erikseen ja suuret arvot yhdistettiin, jotta muuttujissa suuren painon saavat arvot voitaisiin eriyttää.

Aineistossa oli monia muuttujia, joilla oli 7 – 12 arvoa. Näiden arvot korvattiin kirjaimilla, jolloin B-course-ohjelmisto ei ryhmittele arvoja, vaan jokaiselle arvolle saadaan oma todennäköisyys, mikä helpottaa tulosten tulkintaa. Esimerkiksi ohjelma yhdistäisi muuttujan *muu vah* arvot 0 (ei muuta omaisuusvahinkoa) ja 1 (liikenne-merkki), mikä vaikeuttaisi tulosten tarkastelua. Myös muuttujan *tunti* arvot on ryhmitelty viiteen ryhmän, joille annettiin nimet. Tämä helpotti vuorokauden ajan tulkintaa.

Testaamalla havaittiin, että edellä esitetty aineiston muokkaaminen ei vaikuta tuloksina saataviin riippuvuusmalleihin; kummallakin aineistolla saatiin samat riippuvuusmallit. Tulosten tarkastelun yhteydessä muuttujien arvoista käytetään luvussa 4 esiteltyjä merkintöjä.

5.1.3 Muuttujakokonaisuuksiin jakaminen

Tehtyjen rajausten jälkeen jäljelle jäänyt 36 muuttujan aineisto ajettiin B-course-ohjelmistolla kolme kertaa ja tuloksena saatiin jokaisella kerralla sama riippuvuusmalli. Kun useammalla ajokerralla saadaan sama malli, voidaan varmistaa, että B-course-ohjelmiston löytämä paras malli on myös todennäköisin malli (ks. luku 3.2.2). Saadun riippuvuusmallin perusteella aineisto jaettiin kuuteen muuttujakokonaisuuteen (taulukko 5.1) siten, että jokainen riippuvuusmallissa esiintyvä riippuvuussuhde sisältyy vähintään yhteen muuttujakokonaisuuteen. Näissä kuudessa muuttujaryhmässä on 9 – 12 muuttujaa ryhmää kohden, joten niistä saatavia pienempiä riippuvuusmalleja on helpompi tarkastella kuin 36 muuttujan mallia, jossa on huomattavasti enemmän muuttujien välisiä riippuvuussuhteita. 9 – 12 muuttujan riippu-

vuusmalli on myös riittävän pieni, jotta sitä voidaan tutkia käyttämällä Java Playgroud -sovellusta, jonka avulla pystytään tarkastelemaan muuttujien arvojen todennäköisyysjakaumia.

Taulukko 5.1: Aineiston muuttujakokonaisuudet.

| Ryhmä 1 | Ryhmä 2 | Ryhmä 3 |
|----------------------|--------------------|----------------------|
| <i>kp-luokka</i> | <i>asutus</i> | <i>kp-luokka</i> |
| <i>muu vah</i> | <i>kuolleet</i> | <i>muu vah</i> |
| <i>nak-460</i> | <i>loukkaant</i> | <i>nak-150</i> |
| <i>nop-laji</i> | <i>muu vah</i> | <i>nak-300</i> |
| <i>nop-raj</i> | <i>nak-150</i> | <i>nop-vastaan</i> |
| <i>nop-suuntaan</i> | <i>nak-460</i> | <i>onn-paikka</i> |
| <i>nop-vastaan</i> | <i>nop-vastaan</i> | <i>paallyste</i> |
| <i>saa</i> | <i>saa</i> | <i>taaj-merk</i> |
| <i>taaj-merk</i> | <i>taaj-merk</i> | <i>toim-luokka</i> |
| <i>toim-luokka</i> | <i>tien-lev</i> | |
| Ryhmä 4 | Ryhmä 5 | Ryhmä 6 |
| <i>ika</i> | <i>ika</i> | <i>ajor-lkm</i> |
| <i>junction-type</i> | <i>kk</i> | <i>junction-type</i> |
| <i>kulj-alkoh</i> | <i>lampotila</i> | <i>kvl</i> |
| <i>muu vah</i> | <i>muu vah</i> | <i>muu vah</i> |
| <i>nak-150</i> | <i>nak-150</i> | <i>nak-150</i> |
| <i>nak-300</i> | <i>nak-460</i> | <i>nak-460</i> |
| <i>nop-vastaan</i> | <i>nop-raj</i> | <i>nop-vastaan</i> |
| <i>onluokka</i> | <i>nop-vastaan</i> | <i>raskas-kvl</i> |
| <i>ontyyppi</i> | <i>pinta</i> | <i>rist-var</i> |
| <i>osallkm</i> | <i>tunti</i> | <i>tien-lev</i> |
| <i>rask-osall</i> | <i>valoisuus</i> | <i>toim-luokka</i> |
| <i>sukup</i> | <i>vkpv</i> | |

Näitä jokaista kuutta muuttujakokonaisuutta vastaava aineisto ajettiin B-course-ohjelmistolla kolme kertaa. Kukin muuttujakokonaisuus sai jokaisella ajokerrallaan saman riippuvuusmallin, joten kaikkien kokonaisuuksien osalta löydettiin B-course-ohjelmiston toiminnan periaatteiden mukainen paras riippuvuusmalli. Nämä kuusi riippuvuusmallia esitellään seuraavassa alaluvussa.

5.2 Muuttujakokonaisuuksien riippuvuusmallit

Tässä luvussa esitellään lyhyesti, millaisia riippuvuusmalleja tutkittavasta aineistosta löytyy. Mallit esitellään kertomalla muuttujien välisien riippuvuussuhteiden vahvuuksista ja kausaalisuussuhteiden laadusta. B-course-ohjelmisto antaa riippuvuusmallista löytyville riippuvuussuhteille vahvuudet. Jos riippuvuussuhteen vah-

vuus on heikko, kyseisen suhteen poistaminen mallista ei pienennä mallin todennäköisyyttä yhtä paljon kuin vahvan riippuvuussuhteen poistaminen. Riippuvuussuhteiden vahvuudet eivät ole tulosten tulkinnan kannalta yhtä tärkeitä kuin riippuvuussuhteiden laatu. Heikon riippuvuussuhteen voi kuitenkin tarvittaessa jättää vähemmälle huomiolle, jos se ei vaikuta tulosten tarkastelussa kovin tärkeältä.

B-course-ohjelmisto antaa malleista naiivin ja ei-niin-naiivin kausaalisuusmallin. Tässä työssä huomioidaan vain viimeksi mainittu kausaalisuusmalli, sillä se ottaa huomioon myös mallin ulkopuoliset muuttujat ja antaa siten todellisemman kuvan muuttujien välisistä riippuvuuksista. Riippuvuussuhteilla on kolme eri tyyppiä

- suora kausaalisuussuhde, jossa muuttuja A on syy ja muuttuja B seuraus (kuvissa yhtenäinen suunnattu kaari)
- kausaalisuussuhde, jossa muuttuja A on syy ja muuttuja B seuraus TAI muuttujilla A ja B on yhteinen latentti syy (kuvissa katkoviivalla merkitty suunnattu kaari)
- riippuvuussuhde, jonka kausaalisuutta ei tiedetä (kuvissa yhtenäinen suuntaamaton kaari)

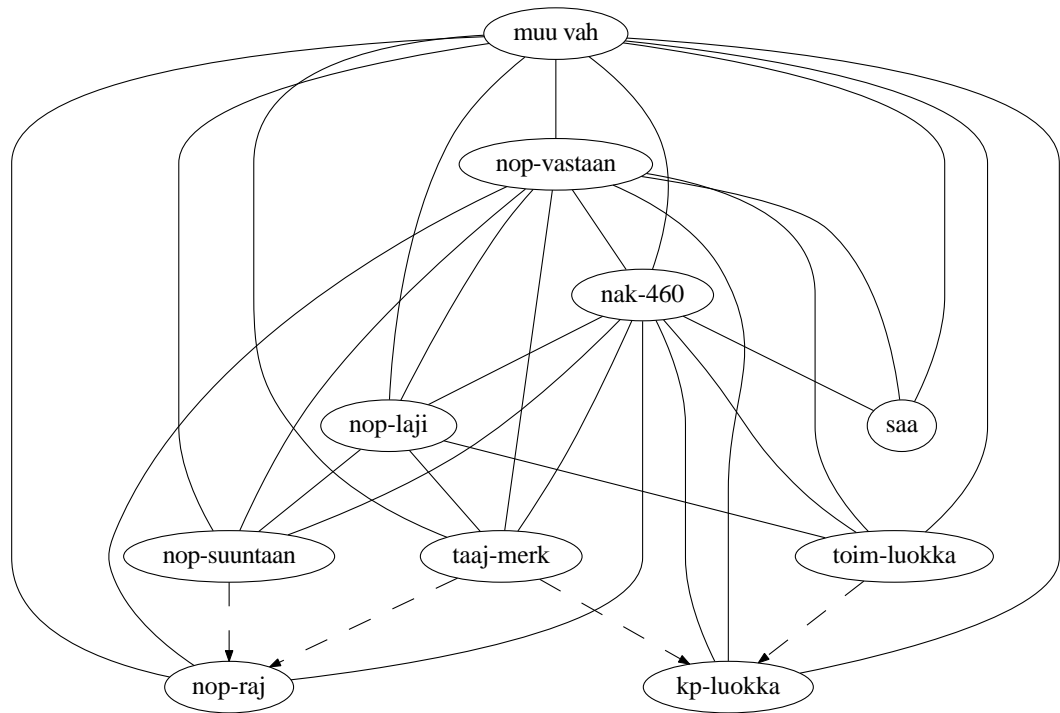
Luvussa 6 analysoidaan tarkemmin malleja muutaman muuttujan tiettyjen arvojen suhteen. Tässä yhteydessä perehdytään myös paremmin riippuvuussuhteiden syihin ja kausaalisuussuhteisiin liittyviin mahdollisiin latentteihin muuttujiin.

5.2.1 Muuttujakokonaisuuden 1 riippuvuusmalli

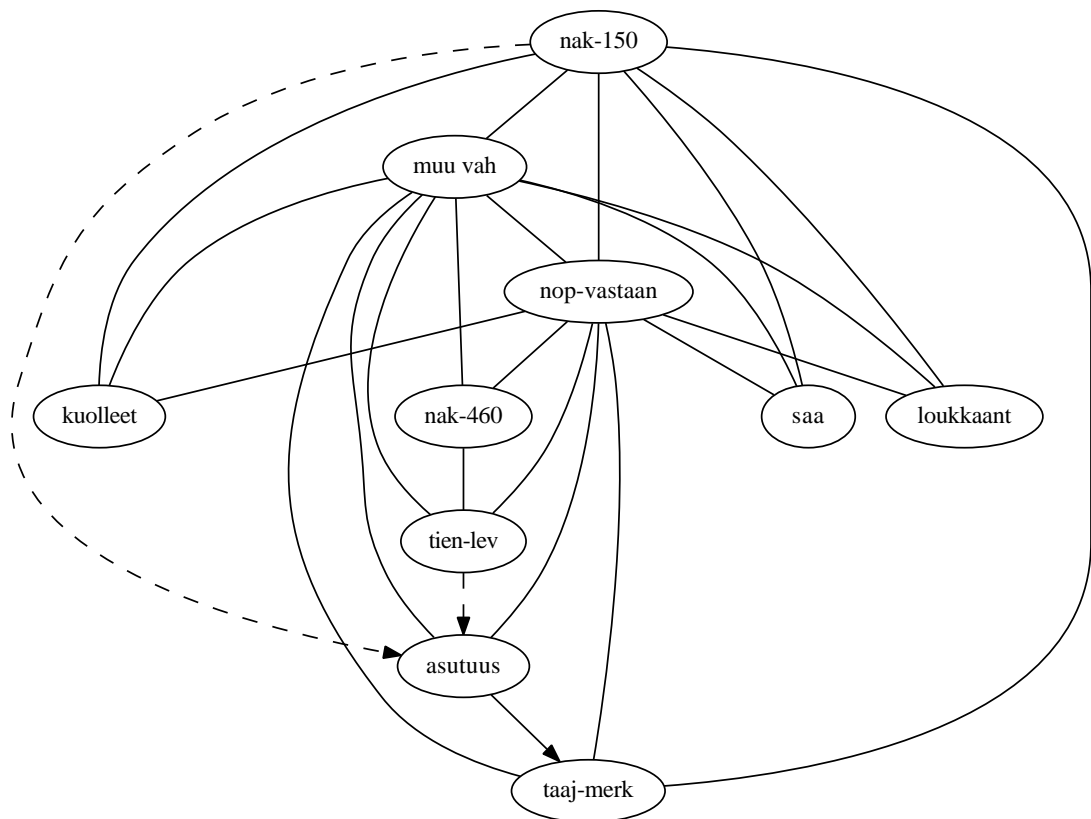
Muuttujakokonaisuuden 1 (ryhmä 1, taulukko 5.1) muuttujien riippuvuusmalli on esitetty kuvassa 5.1(a). Tämän mallin muuttujien väliset riippuvuussuhteet ovat vahvoja tai melko vahvoja lukuun ottamatta muuttujien *taaj-merk* ja *nop-raj* välistä suhdetta.

Muuttujilla *nop-suuntaan* ja *taaj-merk* on kausaalisuussuhde muuttujaan *nop-raj*. B-course-ohjelmiston mukaan nämä suhteet ovat sellaisia, että joko edellä mainitut muuttujat ovat syitä ja jälkimmäinen muuttuja on seuraus, tai muuttujille on olemassa latentti syy. Muuttujien *taaj-merk* ja *toim-luokka* kausaalisuussuhteet muuttujaan *kp-luokka* ovat myös tällaisia.

Edellä esitellyistä kausaalisuussuhteista muuttujien *taaj-merk* ja *nop-raj* välinen suhde voi olla kausaalisuussuhde, jossa *taaj-merk* on syy ja *nop-raj* seuraus, koska taajamissa on pienemmät nopeusrajoitukset kuin taajamien ulkopuolella. Sen sijaan muuttujien *nop-suuntaan* ja *nop-raj* välinen suhde toimisi toiseen suuntaan (*nop-raj* syy ja *nop-suuntaan* seuraus) kausaalisena suhteena, mutta ohjelmiston antamaan suuntaan ei, joten näiden muuttujien kausaalisuussuhteen syynä voisi olla latentti muuttuja. Muuttujien *taaj-merk* ja *kp-luokka* sekä muuttujien *toim-luokka*



(a) Muuttujakokonaisuuden 1 riippuvuusmalli.



(b) Muuttujakokonaisuuden 2 riippuvuusmalli.

Kuva 5.1: Riippuvuusmallit muuttujakokonaisuuksille 1 ja 2. Erilaiset kaaret muuttujien välillä kuvaavat erilaisia muuttujien välisiä riippuvuussuhteita.

ja *kp-luokka* väliset kausaalisuussuhteet voisivat olla syy-seuraussuhteita, sillä yleensä taajamissa teiden talvikunnossapito (*kp-luokka*) toimii paremmin kuin taajamien ulkopuolella, ja toisaalta suurempien teiden (*toim-luokka*) talvikunnossapito toimii tehokkaammin kuin pienempien teiden.

5.2.2 Muuttujakokonaisuuden 2 riippuvuusmalli

Kuvassa 5.1(b) on toisen muuttujakokonaisuuden riippuvuusmalli (ryhmä 2, taulukko 5.1). Mallissa riippuvuussuhteet muuttujien välillä ovat vahvoja tai melko vahvoja, paitsi muuttujien *nak-150* ja *loukkaant* välinen suhde.

Muuttujien *asutus* ja *taaj-merk* välisessä kausaalisessa suhteessa *asutus* on syy ja *taaj-merk* on seuraus. Lisäksi muuttujien *nak-150* ja *tien-lev* riippuvuussuhteet muuttujaan *asutus* ovat sellaisia, joissa edellä mainitut ovat jälkimmäisen syitä, tai muuttujilla on yhteinen latentti syy. Todennäköisesti molemmissa kausaalisuussuhteissa on tausta jokin latentti syy, sillä kumpikaan suhde ei vaikuta suoralta syy-seuraussuhteelta.

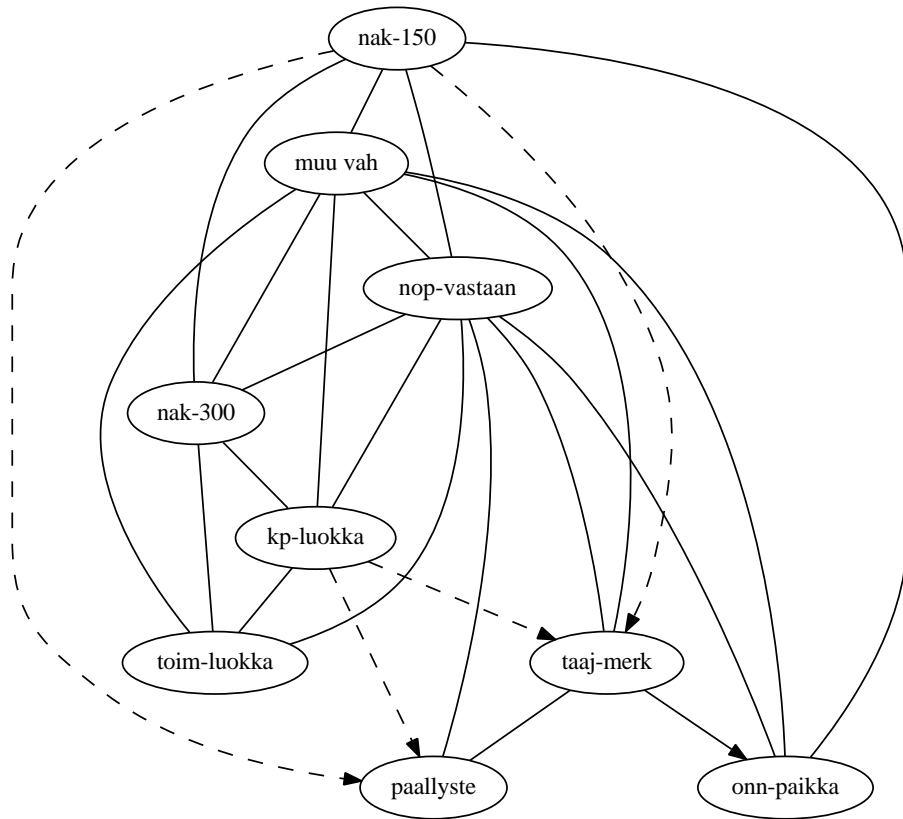
5.2.3 Muuttujakokonaisuuden 3 riippuvuusmalli

Kolmannen muuttujakokonaisuuden (ryhmä 3, taulukko 5.1) riippuvuusmalli on kuvassa 5.2(a). Tämän mallin muuttujien väliset riippuvuussuhteet ovat kaikki vahvoja tai melko vahvoja. B-course-ohjelmiston mukaan mallissa on yksi kausaalisuussuhde, jossa *taaj-merk* on syy ja *onn-paikka* seuraus.

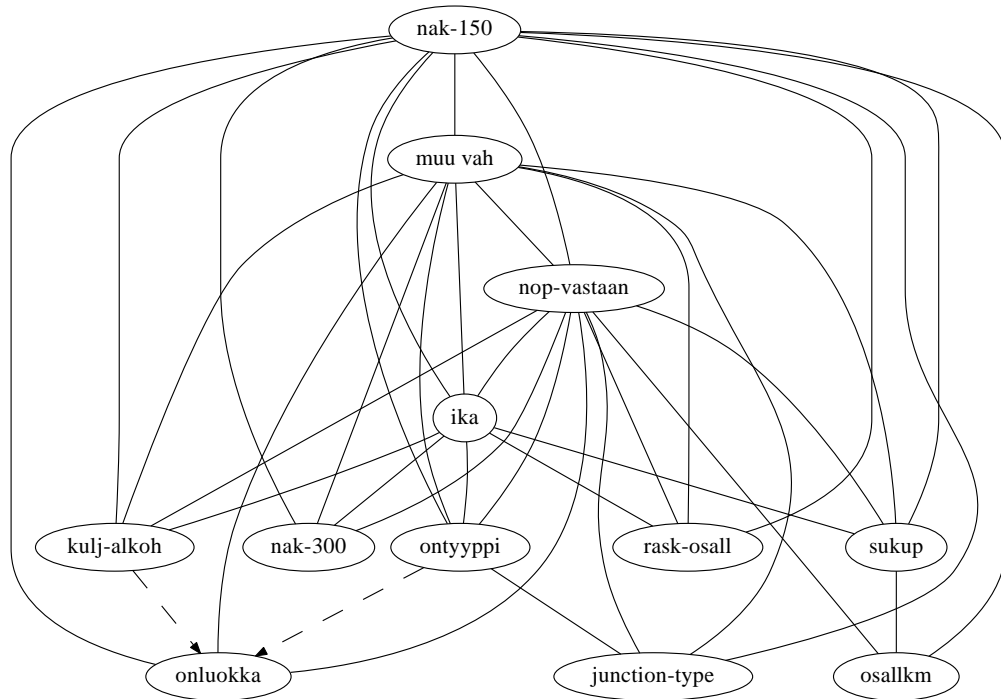
Mallissa on muuttujaparit *nak-150* ja *taaj-merk*; *nak-150* ja *paallyste*; *kp-luokka* ja *taaj-merk* sekä *kp-luokka* ja *paallyste*, joiden välillä on kausaalisuussuhde siten, että ensimmäinen muuttuja on syy ja jälkimmäinen seuraus, tai muuttujilla on latentti syy. Näistä pareista ei yksikään vaikuta selkeältä syy-seurausparilta, joten näillä kausaalisuussuhteille löytyy todennäköisesti latentteja muuttujia selittämään kausaalisuuden muuttujien välillä. Muuttujakokonaisuuden 1 riippuvuusmallissa (kuva 5.1(a)) muuttujilla *kp-luokka* ja *taaj-merk* on samanlainen kausaalisuussuhde kuin tässä mallissa, mutta muuttujien roolit ovat toisin päin, joten näiden muuttujien suhde on melko varmasti latentin muuttujan aiheuttama. Myös muuttujien *nak-150* ja *taaj-merk* välinen riippuvuussuhde löytyy riippuvuusmalleista 1 ja 2, joissa kyseinen suhde ei ole ohjelmiston mukaan kausaalinen.

5.2.4 Muuttujakokonaisuuden 4 riippuvuusmalli

Kuvassa 5.2(b) on neljännen muuttujakokonaisuuden riippuvuusmalli (ryhmä 4, taulukko 5.1). Lähes kaikki muuttujien väliset riippuvuussuhteet mallissa ovat vahvoja tai melko vahvoja; ainoastaan riippuvuussuhde muuttujien *sukup* ja *osallkm* on heikko.



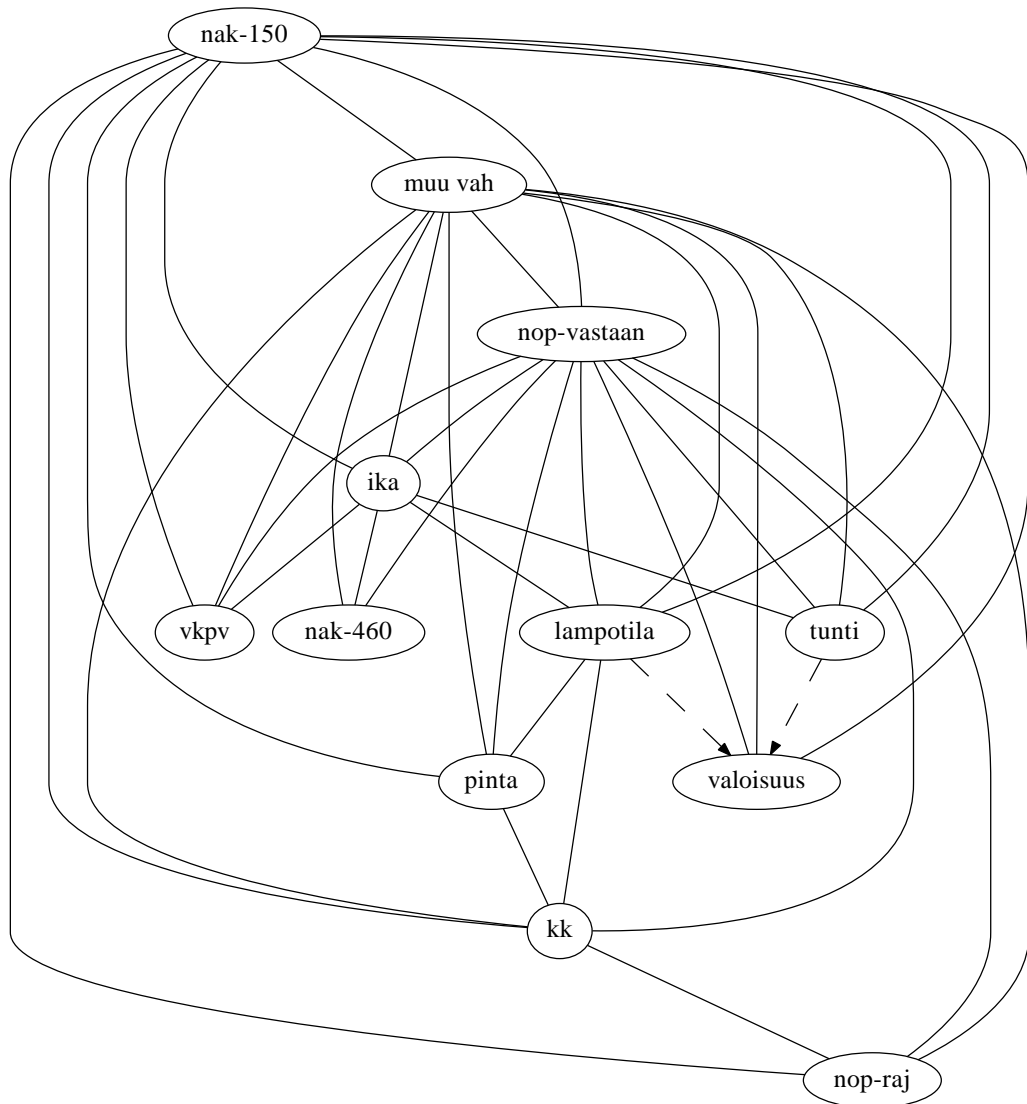
(a) Muuttujakokonaisuuden 3 riippuvuusmalli.



(b) Muuttujakokonaisuuden 4 riippuvuusmalli.

Kuva 5.2: Riippuvuusmallit muuttujakokonaisuuksille 3 ja 4. Erilaiset kaaret muuttujien välillä kuvaavat erilaisia muuttujien välisiä riippuvuussuhteita.

Tässä mallissa vain muuttujien *ontyyppi* ja *onluokka*, sekä muuttujien *kulj-alkoh* ja *onluokka* väleillä on kausaalisuussuhteet, joissa *ontyyppi* ja *kulj-alkoh* ovat syitä ja *onluokka* on seuraus, tai näissä molemmissa kausaalisuussuhteissa on latentti syy. Onnettomuusluokka on määritelty onnettomuustyyppin avulla, joten muuttujien *ontyyppi* ja *onluokka* välinen suhde on syy-seuraussuhde. Sen sijaan muuttujien *kulj-alkoh* ja *onluokka* välinen kausaalinen suhde on todennäköisesti latentin muuttujan aiheuttama.

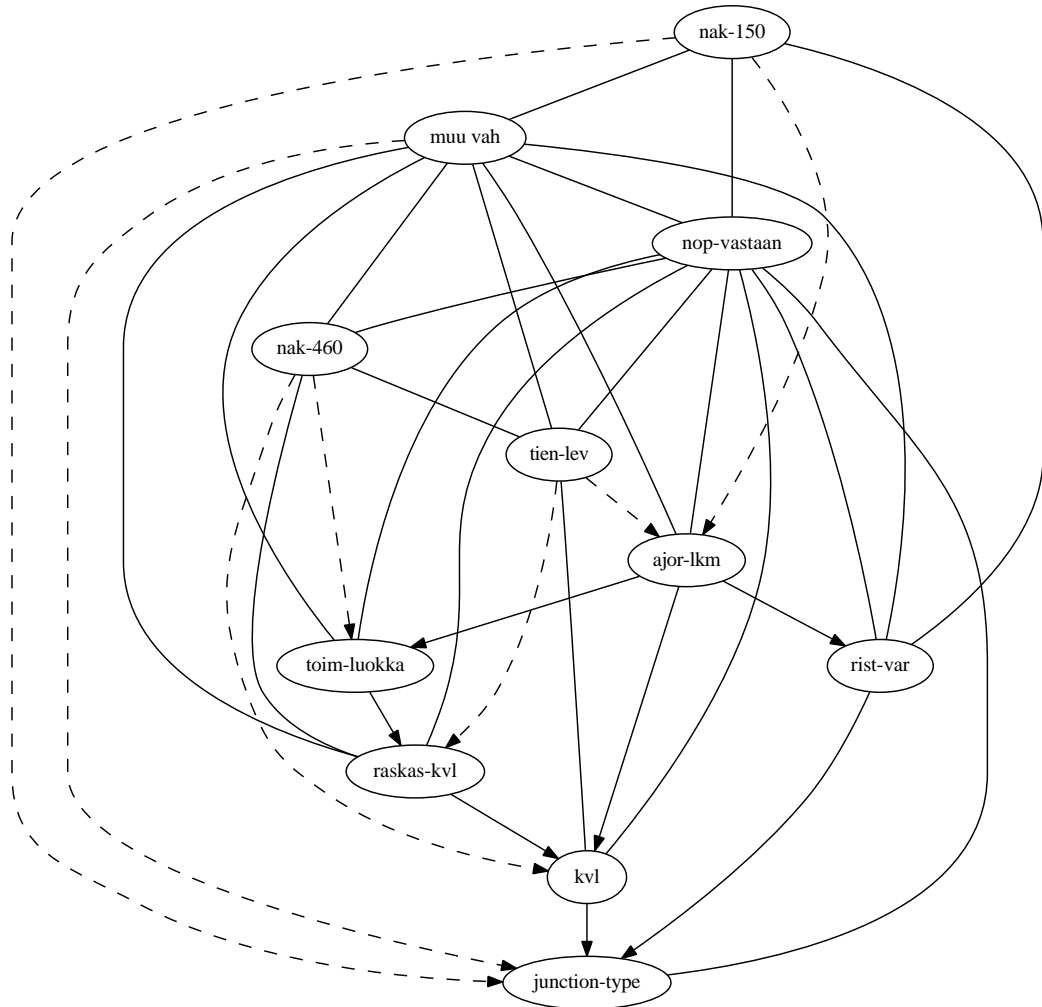


Kuva 5.3: Muuttujakokonaisuuden 5 riippuvuusmalli. Erilaiset kaaret muuttujien välillä kuvaavat erilaisia muuttujien välisiä riippuvuussuhteita.

5.2.5 Muuttujakokonaisuuden 5 riippuvuusmalli

Viidennen muuttujakokonaisuuden (ryhmä 5, taulukko 5.1) riippuvuusmalli on kuvassa 5.3. Kaikki mallin riippuvuussuhteet ovat vahvoja. Muuttujilla *lampotila* ja

tunti on kausaalisuussuhteet muuttujaan *valoisuus* siten, että ensin mainitut ovat syitä ja *valoisuus* on seuraus, tai muuttujilla on latentti syy aiheuttamassa kyseisen suhteen. Muuttujien *lampotila* ja *valoisuus* väliselle kausaalisuussuhteelle voisi olettaa latentiksi syyksi vuodenajan/kuukauden, vuorokaudenajan ja sään yhdistetyn vaikutuksen. Riippuvuusmallissa kuukaudella (*kk*) on suhde muuttujaan *lampotila*, ja vuorokaudenajalla *tunti* on suhde muuttujaan *valoisuus*. säätä vastaavaa muuttujaa mallissa ei ole, joten näiden kolmen tekijän yhteisvaikutusta mallista ei voida päätellä. Muuttujien *tunti* ja *valoisuus* välinen suhde on syy-seuraussuhde.



Kuva 5.4: Muuttujakokonaisuuden 6 riippuvuusmalli. Erilaiset kaaret muuttujien välillä kuvaavat erilaisia muuttujien välisiä riippuvuussuhteita.

5.2.6 Muuttujakokonaisuuden 6 riippuvuusmalli

Ryhmän 6 (taulukko 5.1) muuttujien riippuvuusmalli on esitetty kuvassa 5.4. Mallin kaikki riippuvuussuhteet ovat vahvoja tai melko vahvoja. Muuttujapareilla *ajor-lkm* ja *rist-var*; *ajor-lkm* ja *toim-luokka*; *ajor-lkm* ja *kvl*; *toim-luokka* ja *raskas-kvl*;

raskas-kvl ja *kvl*; *rist-var* ja *junction-type*, sekä *kvl* ja *junction-type* on kausaalisuussuhteet siten, että ensimmäinen muuttuja on syy ja jälkimmäinen seuraus.

Mallin muuttujaparien *nak-150* ja *ajor-lkm*; *nak-150* ja *junction-type*; *muu vah* ja *junction-type*; *nak-460* ja *toim-luokka*; *nak-460* ja *kvl*; *tien-lev* ja *raskas-kvl*, sekä *tien-lev* ja *ajor-lkm* muuttujien välillä on kausaalisuussuhde siten, että ensimmäinen muuttuja on syy ja jälkimmäinen seuraus, tai muuttujilla on latentti syy. Parin *tien-lev* ja *raskas-kvl* välinen suhde on todennäköisesti syy-seuraussuhde, sillä raskaat kulkuneuvot ajavat mieluummin leveillä kuin kapeilla teillä. Muuttujien *tien-lev* ja *ajor-lkm* välillä on selkeän oloinen syy-seuraussuhde, mutta päinvastaiseen suuntaan (*ajor-lkm* syy ja *tien-lev* on seuraus), joten kyseessä voisi olla latentti syy. Muidenkin mainittujen muuttujaparien väliset kausaalisuussuhteet todennäköisesti selittyvät latenteilla muuttujilla.

6. TULOKSIA

Tässä luvussa analysoidaan muiden muuttujien vaikutusta valittuihin muuttujiin. Ensin tutkitaan mitä eroja taajamissa ja taajamien ulkopuolella tapahtuneissa liikenneonnettomuuksissa on. Sen jälkeen tarkastellaan eläinonnettomuuksia ja yli 65-vuotiaille sattuneita liikenneonnettomuuksia. Taulukoiden kokojen rajoittamiseksi tulosten tarkastelussa joidenkin muuttujien arvoja on yhdistetty käsin arvojen samankaltaisuuden perusteella. Lisäksi muuttujien arvot on esitetty selkeyden takia numeroilla, joiden selitykset löytyvät luvusta 4. Niiden muuttujien, joiden arvot on jaettu kolmeen ryhmään, arvot ovat B-course-ohjelmiston ryhmittelemiä.

6.1 Taajamissa ja taajamien ulkopuolella tapahtuneiden liikenneonnettomuuksien erot Pirkanmaalla

Pirkanmaan taajamissa on tapahtunut noin 1300 liikenneonnettomuutta vuosien 2004 – 2008 aikana. Taajamien ulkopuolella sattui lähes 7400 liikenneonnettomuutta samana aikana. Muuttujakokonaisuuksien 1, 2 ja 3 riippuvuusmallien avulla tutkitaan mitä eroja taajamissa ja taajamien ulkopuolella sattuneissa liikenneonnettomuuksissa on. Muuttuja *taaj-merk* on mukana kolmessa riippuvuusmallissa, joten se on koko aineiston kannalta keskeinen muuttuja. Sen avulla on myös helppo jakaa aineiston onnettomuudet kahteen kategoriaan, jotka eroavat merkittävästi toisistaan: taajamissa tapahtuneet onnettomuudet ja taajamien ulkopuolella tapahtuneet onnettomuudet.

Riippuvuusmallissa 1 (kuva 5.1(a)) *taaj-merk* on yhteydessä melkein kaikkiin muihin mallin muuttujiin, ja sillä on kausaalisuussuhde muuttujien *nop-raj* ja *kp-luokka* kanssa. Luvun 5.2.1 mukaisesti *taaj-merk* on kummassakin suhteessa syy ja *nop-raj* ja *kp-luokka* ovat seurauksia. Tässä muuttujakokonaisuudessa *taaj-merk* on siis keskeinen muuttuja.

Riippuvuusmallin 2 kuvasta 5.1(b) nähdään, että *taaj-merk* ei ole erityisen merkittävässä asemassa tässä muuttujakokonaisuudessa. Sillä on samoja riippuvuussuhteita kuin mallissa 1, lisäksi sillä on kausaalinen suhde muuttujan *asutus* kanssa. Muuttuja *asutus* kuvaa asukastiheyttä, ja taajamissa on suurempi asukastiheys kuin niiden ulkopuolella. Tällä suhteella ei ole merkitystä liikenneonnettomuuksien kannalta.

Taulukko 6.1: Muuttujakokonaisuuden 2 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *taaj-merk* arvojen *K* ja *E* suhteen. Muuttujien arvojen selitykset ovat luvussa 4.

| Taajama Ei taajama | | | Taajama Ei taajama | | |
|--------------------|-----|-----|--------------------|-----|-----|
| <i>nak-150</i> | | | <i>nak-460</i> | | |
| arvot | % | % | arvot | % | % |
| 5 – 36 | 0.7 | 1.4 | 0 – 33 | 78 | 31 |
| 37 – 68 | 14 | 4 | 34 – 66 | 21 | 51 |
| 69 – 100 | 85 | 95 | 67 – 100 | 1.2 | 19 |
| <i>muu vah</i> | | | <i>tien-lev</i> | | |
| arvot | % | % | arvot | % | % |
| 0 | 93 | 91 | 35 – 125 | 96 | 85 |
| 1, 2, 7 | 5.7 | 4.3 | 126 – 215 | 3.6 | 11 |
| 3, 4 | 0.9 | 3.9 | 216 – 306 | 0.2 | 4.2 |
| 5, 6 | 0.4 | 0.8 | | | |
| <i>nop-vastaa</i> | | | <i>asutus</i> | | |
| arvot | % | % | arvot | % | % |
| 30 – 60 | 96 | 20 | 0 – 2 | 99 | 34 |
| 61 – 90 | 3.1 | 40 | 3 – 4 | 0.3 | 16 |
| 91 – 120 | 0.6 | 40 | 5 – 6 | 0.3 | 50 |
| <i>loukkaant</i> | | | <i>kuolleet</i> | | |
| arvot | % | % | arvot | % | % |
| 0 | 80 | 84 | 0 | 99 | 99 |
| 1 | 16 | 13 | 1 | 0.6 | 1.3 |
| 2 | 2.5 | 2 | 2 | 0.2 | 0 |
| 3 | 0.4 | 0.6 | ≥ 3 | 0.2 | 0 |
| ≥ 4 | 0.4 | 0.4 | | | |
| <i>saa</i> | | | | | |
| arvot | % | % | | | |
| 1 | 35 | 33 | | | |
| 2 | 44 | 46 | | | |
| 3 | 2 | 2.4 | | | |
| 4 | 12 | 9.9 | | | |
| 5 | 4.2 | 5.9 | | | |
| 6,7 | 3.1 | 3.4 | | | |

Kolmannessa riippuvuusmallissa (kuva 5.2(a)) muuttujalla *taaj-merk* on riippuvuussuhteita lähes kaikkien muiden mallin muuttujien kanssa. Sillä on kausaalisuussuhde muuttujan *onn-paikka* kanssa, jossa *taaj-merk* on syy ja *onn-paikka* seuraus. Taajamissa ja taajamien ulkopuolella on erilaisia onnettomuuspaikkoja, joten tämä kausaalisuussuhde helposti pääteltävissä. Lisäksi muuttujalla *taaj-merk* on kausaalisuussuhteet muuttujien *nak-150* ja *kp-luokka* kanssa. Kuten luvussa 5.2.3 todetaan, muuttujien *nak-150* ja *taaj-merk* kausaalisuussuhteelle syynä voisi olla jokin latent-

Taulukko 6.2: Muuttujakokonaisuuden 1 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *taaj-merk* arvojen *K* ja *E* suhteen. Taulukosta puuttuvat muuttujat *muu vah*, *nop-vastaan*, *nak-460* ja *saa*, koska niiden jakaumat löytyvät taulukosta 6.1. Muuttujien arvojen selitykset ovat luvussa 4.

| Taajama | | | Ei taajama | | |
|---------------------|-----|-----|------------------|-----|-----|
| <i>nop-suuntaan</i> | | | <i>nop-raji</i> | | |
| arvot | % | % | arvot | % | % |
| 30 – 60 | 97 | 20 | 30 – 60 | 99 | 24 |
| 61 – 90 | 2.9 | 38 | 61 – 90 | 0.3 | 53 |
| 91 – 120 | 0.3 | 41 | 91 – 120 | 0.3 | 23 |
| <i>nop-laji</i> | | | <i>kp-luokka</i> | | |
| arvot | % | % | arvot | % | % |
| 1 | 5.5 | 81 | 1 | 2.2 | 39 |
| 2 | 0.3 | 4.5 | 2 | 10 | 21 |
| 3 | 0 | 0.2 | 3 | 43 | 28 |
| 4 | 37 | 13 | 4 | 33 | 0.7 |
| 5 | 20 | 0.3 | 5 | 11 | 11 |
| 6 | 37 | 0 | 6 | 0.3 | 0.5 |
| <i>toim-luokka</i> | | | | | |
| arvot | % | % | | | |
| 1 | 2.2 | 47 | | | |
| 2 | 3.2 | 10 | | | |
| 3 | 43 | 28 | | | |
| 4 | 52 | 15 | | | |

timuuttuja, ja muuttujien *kp-luokka* ja *taaj-merk* kausaalinen suhde selittyisi paremmin niin päin, että *taaj-merk* olisi syy ja *kp-luokka* seuraus. B-course-ohjelmisto tarjoaa kuitenkin kyseiselle kausaalisuussuhteelle päinvastaista syy-seuraussuhdetta tai latenttia syytä.

Vain noin joka viidennessä liikenneonnettomuudessa loukkaantuu ihmisiä ja vain yhdessä sadasta onnettomuudesta kuolee ihminen (*loukkaant* ja *kuolleet*, taulukko 6.1). Taulukon mukaan taajama-alueen onnettomuuksissa loukkaantuu hieman todennäköisemmin ihmisiä kuin taajaman ulkopuolella sattuneissa onnettomuuksissa. Liikenneonnettomuuksissa kuolleitten osalta eroa ei juurikaan ole. Taajama-alueen onnettomuuksissa loukkaantumistodennäköisyyttä nostaa se, että taajamissa kävelijät ja pyöräilijät joutuvat useammin liikenneonnettomuuteen kuin taajamien ulkopuolella.

Taajamien onnettomuuspaikoissa näkyvyys on heikompi kuin taajamien ulkopuolella. Tämä voidaan päätellä muuttujien *nak-150*, *nak-300* ja *nak-460* arvojen todennäköisyysjakaumista (taulukot 6.1 ja 6.3). Etenkin pidempien (300 m ja 460 m) matkojen osalta näkemäprosentit taajamissa ovat suurimmassa osassa onnettomuuksia olleet pieniä. Taulukoista 6.1 ja 6.2 havaitaan selkeät erot myös ajonopeuk-

Taulukko 6.3: Muuttujakokonaisuuden 3 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *taaj-merk* arvojen *K* ja *E* suhteen. Taulukosta puuttuu muuttujat *nak-150*, *muu vah*, *nop-vastaan*, *kp-luokka* ja *toim-luokka*, koska niiden jakaumat löytyvät taulukoista 6.1 ja 6.2. Muuttujan *onn-paikka* arvo 7 ei esiinny aineistossa kertaakaan. Muuttujien arvojen selitykset ovat luvussa 4.

| Taajama Ei taajama | | | Taajama Ei taajama | | |
|--------------------|-----|-----|--------------------|-----|-----|
| <i>paallyste</i> | | | <i>onn-paikka</i> | | |
| arvot | % | % | arvot | % | % |
| 1 | 91 | 88 | 1 | 82 | 97 |
| 2 | 5.9 | 11 | 2, 3 | 11 | 0.3 |
| 3, 4 | 2.1 | 0.9 | 4, 5 | 1.1 | 0 |
| 5, 6 | 0.9 | 0.2 | 6 (7) | 1.0 | 1.0 |
| | | | 8 | 3.4 | 1.8 |
| | | | 9 | 1.0 | 0.1 |
| <i>nak-300</i> | | | | | |
| arvot | % | % | | | |
| 0 – 33 | 36 | 10 | | | |
| 34 – 66 | 45 | 18 | | | |
| 67 – 100 | 19 | 72 | | | |

sisä (*nop-suuntaan* ja *nop-vastaan*) sekä nopeusrajoituksissa (*nop-raj*). Taajamis- sa nopeusrajoitus on lähes aina korkeintaan 60 km/h, myös ajonopeudet onnettomuustilanteissa ovat melkein aina olleet alle 60 km/h. Taajamien ulkopuolella nopeusrajoitukset ja ajonopeudet sen sijaan ovat pääasiassa olleet yli 60 km/h. Taajamien ulkopuolella nopeusrajoituksen tyyppi (*nop-laji*) on pääasiassa tiekohtainen nopeusrajoitus, kun taas taajamissa nopeusrajoitus on yleensä paikallinen rajoitus, nopeusrajoitusalueen rajoitus tai taajaman yleisrajoitus (taulukot 6.2 ja 4.7).

Tyypillisiä eroja taajamissa ja taajaman ulkopuolella sattuneissa onnettomuuksissa on myös erot tieluokassa (*toim-luokka*) ja talvihoitoluokassa (*kp-luokka*). Taajamissa lähes kaikki liikenneonnettomuudet ovat tapahtuneet seutu- tai yhdysteillä, jotka ovat taajamien yleisemmät tieluokat. Taajamien ulkopuolella sitä vastoin yli puolet onnettomuuksista on tapahtunut valtateillä (taulukko 6.2 ja luku 4.3). Valtatiet ovat muihin taajaman ulkopuolisiin teihin verrattuna yleensä vilkkaammin liikennöityjä kuin muut tiet, ja siten niillä tapahtuu muita teitä enemmän onnettomuuksia. Talvihoitoluokkaan vaikuttavat tien liikennemäärä, liikenteen luonne ja koostumus, sekä nopeusrajoitus. Näistä syistä johtuen valtatie ja suuri osa kantateista kuuluu hoitoluokkaan 1 (1s, 1, 1b ja T1b). Taulukosta 6.2 nähdään, että vain reilu kymmenes osa sekä taajamissa että taajamien ulkopuolella sattuneista onnettomuuksista on tapahtunut hoitoluokan II tai III (arvot 5 ja 6, taulukko 4.2) teillä. Noin 90 prosenttia liikenneonnettomuuksista tapahtuu siis parempaan hoitoluokkaan kuuluvilla teillä. Tärkeimpänä syynä tähän on näiden teiden suuremmat

liikennemäärät.

Taajamien ulkopuolella lähes kaikki liikenneonnettomuudet ovat tapahtuneet ajoradalla (arvo 1) (*onn-paikka*, taulukot 6.3, 4.8). Lisäksi muutamia onnettomuuksia on tapahtunut silloilla ja eritasoliittymien rampeissa (arvot 6 ja 8). Taajamissakin suurin osa onnettomuuksista on tapahtunut ajoradalla. Myös suojatie- ja kevyen liikenteen väylän onnettomuuksia on yli 10 prosenttia (arvot 2 – 3), näistä suurin osa on ollut suojatieonnettomuuksia. Loput taajama-alueen onnettomuudet ovat jakautuneet melko tasaisesti muihin (taulukossa 4.8 lueteltuihin) paikkoihin. Ero taajamanulkopuolisten ja taajamassa tapahtuneiden onnettomuuksien välillä johtuu siitä, että taajamissa on enemmän jalankulkijoita ja pyöräilijöitä, sekä enemmän esimerkiksi suojateitä ja kevyen liikenteen väyliä.

6.2 Eläinonnettomuudet Pirkanmaalla

Pirkanmaalla on vuosien 2004 – 2008 aikana tapahtunut noin 3200 eläinonnettomuutta, joista hirvi- ja peuraonnettomuuksia on noin 3100. Näitä onnettomuuksia tarkastellaan ensin muuttujakokonaisuuden 4 riippuvuusmallin (kuva 5.2(b)) avulla. Sen jälkeen kootaan muutamia aiheen kannalta kiinnostavia muuttujia ja tarkastellaan niiden yhteyttä eläinonnettomuuksiin.

Eläinonnettomuudet ovat muuttujan *onluokka* arvot 10 – 12. Riippuvuusmallista (kuva 5.2(b)) nähdään, että muuttujalla *onluokka* on riippuvuussuhde muuttujien *kulj-alkoh*, *ontyyppi*, *nak-150*, *muu vah* ja *nop-vastaan* kanssa. Näiden lisäksi myös *ika*-muuttuja riippuvuussuhteidensa perusteella tärkeä muuttujan *onluokka* suhteen.

Taulukossa 6.4 on esitetty muuttujakokonaisuuden 4 muuttujien arvojen todennäköisyysjakaumat eläinonnettomuuksien osalta sekä koko aineistossa. Verrattaessa eläinonnettomuuksia kaikkiin liikenneonnettomuuksiin nähdään, että kuljettajan sukupuolella (*sukup*) ei ole vaikutusta eläinonnettomuuksiin; muuttujan todennäköisyysjakaumat ovat samanlaiset kummassakin onnettomuusryhmässä. Kuljettajan osalta eroa jakaumista löytyy muuttujien *kulj-alkoh* ja *ika* kohdalla. Eläinonnettomuuksissa kuljettaja ei ole kertaakaan ollut alkoholin vaikutuksen alaisena. 36 – 64-vuotiaiden kuljettajien osuus eläinonnettomuuksissa on hieman suurempi kuin kaikissa onnettomuuksissa.

Onnettomuustyyppin (*ontyyppi*) arvojen todennäköisyysjakaumien erot taulukossa 6.4 johtuvat siitä, että onnettomuusluokat on määriteltä onnettomuustyyppien avulla. Muuttujan *ontyyppi* arvo 90 vastaa eläinonnettomuutta (liite A). Eläinonnettomuudet ovat yleensä sellaisia, että ajoneuvo törmää eläimeen, eikä onnettomuudessa ole muita osapuolia. Tämä näkyy myös osallisten lukumäärissä (*osallkm*); lähes kaikissa eläinonnettomuuksissa on ollut kaksi osallista, ajoneuvo ja eläin. Eläinonnettomuuksissa ei juuri koskaan aiheudu muuta omaisuusvahinkoa (*muu vah*).

Taulukko 6.4: Muuttujakokonaisuuden 4 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *onluokka* arvojen 10 – 12 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *onluokka* ei ole valittu arvoa. Muuttujien arvojen selitykset ovat luvussa 4.

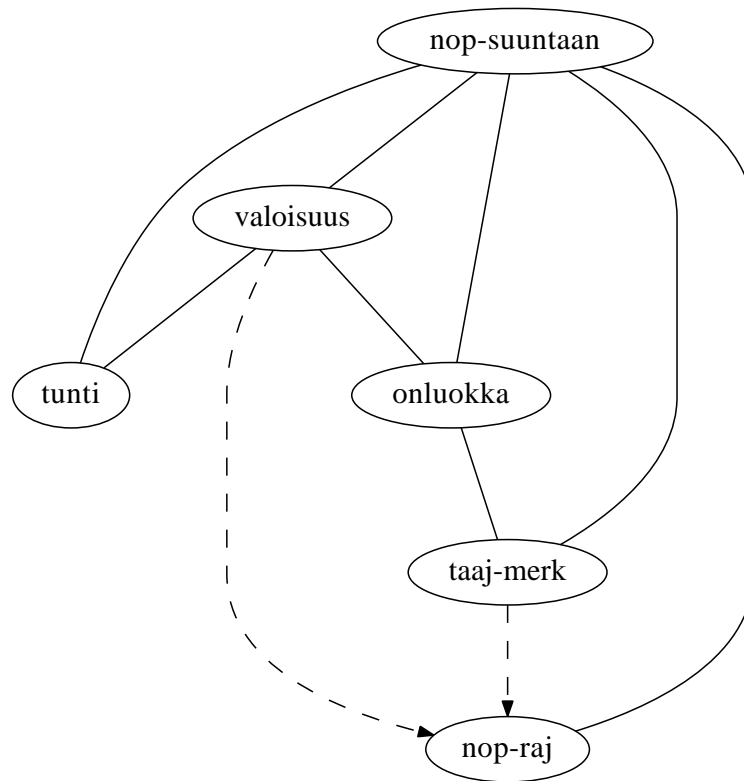
| Eläinonnettomuus | | | Eläinonnettomuus | | |
|--------------------|-----|-----|----------------------|-----|-----|
| <i>nak-150</i> | | | <i>ontyyppi</i> | | |
| arvot | % | % | arvot | % | % |
| 5 – 36 | 0.1 | 1.3 | 0 – 33 | 0 | 26 |
| 37 – 68 | 3.3 | 5.4 | 34 – 66 | 0 | 13 |
| 69 – 100 | 97 | 93 | 67 – 99 | 100 | 61 |
| <i>muu vah</i> | | | <i>ika</i> | | |
| arvot | % | % | arvot | % | % |
| 0 | 99 | 91 | 6 – 35 | 33 | 39 |
| 1, 2, 7 | 0.4 | 4.4 | 36 – 64 | 63 | 53 |
| 3, 4 | 0.2 | 3.4 | 65 – 93 | 4.7 | 8.1 |
| 5, 6 | 0 | 0.7 | | | |
| <i>nop-vastaan</i> | | | <i>junction-type</i> | | |
| arvot | % | % | arvot | % | % |
| 30 – 60 | 13 | 31 | 0 – 3 | 3.5 | 16 |
| 61 – 90 | 34 | 35 | 4 – 6 | 0.9 | 4.1 |
| 91 – 120 | 53 | 34 | 7 – 10 | 96 | 80 |
| <i>kulj-alkoh</i> | | | <i>sukup</i> | | |
| arvot | % | % | arvot | % | % |
| Ei | 100 | 93 | M | 73 | 73 |
| Kyllä | 0 | 7 | N | 27 | 27 |
| <i>nak-300</i> | | | <i>osallkm</i> | | |
| arvot | % | % | arvot | % | % |
| 0 – 33 | 7.2 | 14 | 1 | 0 | 22 |
| 34 – 66 | 15 | 22 | 2 | 97 | 72 |
| 67 – 100 | 78 | 64 | ≥ 3 | 2.5 | 6 |
| <i>rask-osall</i> | | | | | |
| arvot | % | % | | | |
| Ei | 83 | 84 | | | |
| On | 17 | 16 | | | |

Tämä johtuu siitä, että suurin osa eläinonnettomuuksista tapahtuu paikoissa, joissa ei ole taulukossa 4.4 lueteltuja vahingonkohteita.

Useimmiten eläinonnettomuudet tapahtuvat ”suorilla” tieosuuksilla, kuten muuttujan *junction-type* arvojen todennäköisyysjakaumasta nähdään. Eläinonnettomuuksien osalta arvoilla 0 – 3 on selkeästi pienempi todennäköisyys ja arvoilla 7 – 10 suurempi todennäköisyys kuin vastaavilla arvoilla kaikkien onnettomuuksien osalta. Arvot 0 – 3 vastaavat risteyksiä ja kiertoliittymiä kun taas arvot 7 – 10 liittyvät ”suoriin” tieosuuksiin (taulukko 4.9).

Tarkasteltaessa muuttujien *nak-150* ja *nak-300* arvoja taulukossa 6.4 huomataan, että molemmat muuttujat saavat eläinonnettomuuksien osalta suurilla arvoilla suuremmat todennäköisyydet kuin kaikkien onnettomuuksien todennäköisyysjakaumassa. Näiden lukemien perusteella voidaan päätellä, että suuri osa tarkastelun kohteena olevista onnettomuuksista on tapahtunut aineistoon nähden keskimääräistä suuremmilla teillä ja/tai näköesteettömämmillä alueilla, koska näkemäprosentit ovat suuria. Luvussa 6.1 esiteltyjen tulosten perusteella voidaan olettaa, että suuri osa eläinonnettomuuksista tapahtuu taajamien ulkopuolella. Luvun 6.1 perusteella myös muuttujan *nop-vastaan* arvojen todennäköisyydet tukevat tätä oletusta: vastaantulijoiden ajonopeuksien 91 – 120 km/h todennäköisyys on huomattavasti suurempi eläinonnettomuuksissa kuin kaikissa onnettomuuksissa.

Tulosten tarkastelun yhteydessä kiinnostaviksi muuttujiksi nousivat *taaj-merk*, *nop-raaj* ja *nop-suuntaan*. Lisäksi, koska tiedetään, että suurin osa hirvi- ja peurakolareista tapahtuu hämärän aikaan, otetaan tarkasteluun mukaan myös muuttujat *valoisuus* ja *tunti*. Näiden muuttujien riippuvuussuhteet *onluokka*-muuttujaan on esitetty kuvassa 6.1. Koska muuttujaan *onluokka* sisältyvät kaikki onnettomuusluokat, eikä ainoastaan eläinonnettomuudet, mallin riippuvuussuhteita ei voida tulkita vain eläinonnettomuuksien osalta. Tämän takia näiden muuttujien osalta tarkastelussa keskitytään taulukon 6.5 todennäköisyysjakaumiin.



Kuva 6.1: Muuttujien *taaj-merk*, *nop-raaj*, *nop-suuntaan*, *valoisuus* ja *tunti* riippuvuussuhteet toisiinsa sekä muuttujaan *onluokka*.

Taulukko 6.5: Muuttujien *nop-suuntaan*, *nop-raj*, *valoisuus*, *tunti* ja *taaj-merk* arvojen todennäköisyysjakaumat muuttujan *onluokka* arvojen 10 – 12 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *onluokka* ei ole valittu arvoa. Muuttujien arvojen selitykset ovat luvussa 4.

| Eläinonnettomuus | | | Eläinonnettomuus | | |
|---------------------|-----|-----|------------------|----|-----|
| <i>nop-suuntaan</i> | | | <i>nop-raj</i> | | |
| arvot | % | % | arvot | % | % |
| 30 – 60 | 14 | 33 | 30 – 60 | 17 | 37 |
| 61 – 90 | 32 | 32 | 61 – 90 | 53 | 42 |
| 91 – 120 | 54 | 36 | 91 – 120 | 30 | 21 |
| <i>valoisuus</i> | | | <i>tunti</i> | | |
| arvot | % | % | arvot | % | % |
| 1 | 28 | 53 | 0 – 4 | 14 | 9.7 |
| 2 | 17 | 9.6 | 5 – 9 | 20 | 21 |
| 3 | 46 | 23 | 10 – 14 | 12 | 23 |
| 4 | 8.2 | 14 | 15 – 18 | 26 | 27 |
| | | | 19 – 23 | 27 | 20 |
| <i>taaj-merk</i> | | | | | |
| arvot | % | % | | | |
| K | 1.9 | 19 | | | |
| E | 98 | 81 | | | |

Taulukosta 6.5 nähdään vahvistus sille, että lähes kaikki eläinonnettomuudet ovat tapahtuneet taajamien ulkopuolella (*taaj-merk*). Eläinonnettomuuksissa ajonopeudet (*nop-suuntaan*) ovat olleet useammin suurempia kuin kaikissa onnettomuuksissa, mikä johtuu varmaankin siitä, että ajonopeudet taajamien ulkopuolella ovat yleensä suurempia kuin taajamissa. Toisaalta suurella nopeudella ajettaessa kuljettajalle jää vähemmän aikaa reagoida yllättäen tielle ilmestyneeseen eläimeen. Yli puolet onnettomuuksista on tapahtunut teillä, joilla nopeusrajoitus on 61 – 90 km/h. Riista-aidat todennäköisesti pienentävät sellaisten teiden osuutta eläinonnettomuustilastossa, joiden nopeusrajoitus on yli 90 km/h.

Muuttujan *valoisuus* arvojen todennäköisyysjakaumasta nähdään, että 63 prosenttia eläinonnettomuuksista on tapahtunut hämärässä tai pimeässä (taulukko 6.5, arvot 2 ja 3) kun taas kaikista onnettomuuksista yli puolet on tapahtunut valoisaan aikaan (arvo 1). Lisäksi muuttujan *tunti* arvojen todennäköisyysjakauman perusteella voidaan todeta, että eläinonnettomuuksia tapahtuu muita onnettomuuksia todennäköisemmin iltaseitsemän ja aamuneljän välisenä aikana. Syynä tähän on se, että eläimet liikkuvat enemmän hämärän ja pimeän aikaan, jolloin myös autoilijan on vaikeampi huomata niitä ajoissa.

6.3 Yli 65-vuotiaiden aiheuttamat liikenneonnettomuudet Pirkanmaalla

Vuosien 2004 – 2008 aikana Pirkanmaalla on sattunut lähes 600 liikenneonnettomuutta, jotka ovat aiheutuneet yli 65-vuotiaista kuljettajista. Tässä luvussa selvitetään, mitä eroja on yli 65-vuotiaiden kuljettajien ajamalla liikenneonnettomuuksilla ja kaikenikäisten ajamalla liikenneonnettomuuksilla. Muuttuja *ika* kuuluu muuttujakokonaisuuksiin 4 ja 5. Yli 65-vuotiaiden ajamia onnettomuuksia tutkitaan ensin näiden muuttujakokonaisuuksien riippuvuusmallien (kuvat 5.2(b) ja 5.3) avulla. Tulosten tarkastelun yhteydessä kootaan muita esiin nousevia muuttujia. Lopuksi tutkitaan näiden muuttujien yhteyttä *ika*-muuttujaan.

Kuvasta 5.2(b) nähdään, että *ika*-muuttuja on hyvin keskeisessä asemassa riippuvuusmallissa 4, joten kaikki mallin muuttujat ovat merkityksellisiä sen suhteen. Riippuvuusmallissa 5 (kuva 5.3) muuttujalla *ika* on riippuvuussuhde muuttujien *nak-150*, *muu vah*, *nop-vastaan*, *vkpv*, *nak-460*, *lampotila* ja *tunti* kanssa, eli tässäkin riippuvuusmallissa *ika* on keskeinen muuttuja.

Muuttujan *sukup* arvoista taulukossa 6.6 huomataan, että yli 65-vuotiaiden osalta miesten osuus liikenneonnettomuuksissa on isompi kuin kaikenikäisten tilastossa. Tämä johtuu todennäköisesti siitä, että yli 65-vuotiaissa on suhteellisesti enemmän autoilevia miehiä kuin koko väestössä, ja siten myös miesten osuus tämän ikäryhmän onnettomuustilastossa on suurempi. Muuttujan *kulj-alkoh* todennäköisyysjakaumista nähdään, että yli 65-vuotiaiden ajamissa liikenneonnettomuuksissa on hieman harvemmin alkoholilla vaikutusta asiaan kuin kaikenikäisten onnettomuuksissa.

Tarkasteltaessa muuttujien *onluokka* ja *ontyyppi* arvojen todennäköisyysjakauksia taulukossa 6.6 huomataan, että onnettomuustyypeillä 34 – 66 ja onnettomuusluokilla 1 – 5 on yli 65-vuotiaiden jakaumassa suuremmat osuudet kuin kaikenikäisten jakaumassa. Toisaalta taas onnettomuustyypeillä 67 – 99 ja onnettomuusluokilla 7 – 10 yli 65-vuotiaiden jakaumassa pienemmät arvot kuin kaikenikäisten jakaumassa. Näistä lukemista voidaan päätellä, että yli 65-vuotiaille sattuu liikenneonnettomuuksia risteysalueilla muita ikäryhmiä useammin, mutta muita ikäryhmiä vähemmän jalankulkija-, pyöräilijä- ja eläinonnettomuuksia. Muuttujan *junction-type* todennäköisyysjakaumat vahvistavat nämä päätelmät: arvoilla 0 – 3 (sisältää risteykset ja kiertoliittymät, taulukko 4.9) yli 65-vuotiaille on suurempi todennäköisyys kuin kaikilla ikäryhmillä. Yli 65-vuotiaiden risteysalueiden onnettomuudet johtuvat todennäköisesti iän myötä heikentyneistä havaintokyvystä ja reaktionopeudesta. Kuitenkin suurin osa yli 65-vuotiaille sattuneista onnettomuuksista tapahtuu *junction-type*-muuttujan jakauman mukaan muualla kuin risteysalueilla. Omaisuusvahinkojen suurempi osuus johtuu todennäköisesti siitä, että muihin ikäryhmiin verrattuna vanhoille ihmisille sattuu enemmän risteysalueonnettomuuksia.

Taulukko 6.6: Muuttujakokonaisuuden 4 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *ika* arvojen 65 – 93 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *ika* ei ole valittu arvoa. Muuttujien arvojen selitykset ovat luvussa 4.

| Yli 65-vuotiaat | | | Yli 65-vuotiaat | | |
|-------------------|-----|-----|----------------------|-----|-----|
| <i>nak-150</i> | | | <i>ontyyppi</i> | | |
| arvot | % | % | arvot | % | % |
| 5 – 36 | 0.8 | 1.3 | 0 – 33 | 27 | 26 |
| 37 – 68 | 5.4 | 5.4 | 34 – 66 | 26 | 13 |
| 69 – 100 | 94 | 93 | 67 – 99 | 47 | 61 |
| <i>muu vah</i> | | | <i>onluokka</i> | | |
| arvot | % | % | arvot | % | % |
| 0 | 88 | 91 | 1 – 5 | 55 | 49 |
| 1, 2, 7 | 7.1 | 4.4 | 6 – 9 | 14 | 12 |
| 3, 4 | 4.9 | 3.4 | 10 – 13 | 30 | 39 |
| 5, 6 | 0.1 | 0.7 | | | |
| <i>nop-vastaa</i> | | | <i>junction-type</i> | | |
| arvot | % | % | arvot | % | % |
| 30 – 60 | 38 | 31 | 0 – 3 | 23 | 16 |
| 61 – 90 | 38 | 35 | 4 – 6 | 5.7 | 4.1 |
| 91 – 120 | 24 | 34 | 7 – 10 | 72 | 80 |
| <i>kulj-alkoh</i> | | | <i>sukup</i> | | |
| arvot | % | % | arvot | % | % |
| Ei | 98 | 93 | M | 83 | 73 |
| Kyllä | 2.4 | 7 | N | 17 | 27 |
| <i>nak-300</i> | | | <i>osallkm</i> | | |
| arvot | % | % | arvot | % | % |
| 0 – 33 | 13 | 14 | 1 | 18 | 22 |
| 34 – 66 | 28 | 22 | 2 | 76 | 72 |
| 67 – 100 | 59 | 64 | ≥ 3 | 6.4 | 6 |
| <i>rask-osall</i> | | | | | |
| arvot | % | % | | | |
| Ei | 89 | 84 | | | |
| On | 11 | 16 | | | |

Yli 65-vuotiaille sattuu kaikenikäisiä useammin liikenneonnettomuuksia, joissa on kaksi osallista, mutta hieman harvemmin sellaisia, joissa on raskas kulkuneuvo mukana (muuttujat *osallkm* ja *rask-osall*, taulukko 6.6). Ikäihmisten liikenneonnettomuuksia aiheutuu hieman useammin muuta omaisuusvahinkoa (muu vah, taulukko 6.1) kuin muissa liikenneonnettomuuksissa.

Luvun 6.1 tuloksien perusteella nopeusrajoituksista (*nop-raj*), vastaantulijoitten ajonopeuksista (*nop-vastaa*) ja näkemäprosenttien (*nak-150*, *nak-300* ja *nak-460*) (taulukot 6.6 ja 6.7) jakaumista voidaan päätellä, että yli 65-vuotiaille sattuu liiken-

Taulukko 6.7: Muuttujakokonaisuuden 5 riippuvuusmallin muuttujien arvojen todennäköisyysjakaumat muuttujan *ika* arvojen 65 – 93 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *ika* ei ole valittu arvoa. Taulukosta puuttuvat muuttujat *nak-150*, *muu vah* ja *nop-vastaan*, koska niiden jakaumat löytyvät taulukosta 6.6. Muuttujien arvojen selitykset ovat luvussa 4.

| Yli 65-vuotiaat | | | Yli 65-vuotiaat | | |
|-----------------|-----|-----|------------------|------|------|
| <i>pinta</i> | | | <i>tunti</i> | | |
| arvot | % | % | arvot | % | % |
| 1,2 | 72 | 71 | 0 – 4 | 2.7 | 8.7 |
| 3 | 0.6 | 0.5 | 5 – 9 | 15 | 21 |
| 4 | 5.8 | 6.7 | 10 – 14 | 42 | 23 |
| 5 | 4.5 | 4.4 | 15 – 18 | 31 | 27 |
| 6 | 15 | 14 | 19 – 23 | 9.7 | 21 |
| 7 | 3 | 3.1 | | | |
| <i>lampoila</i> | | | <i>valoisuus</i> | | |
| arvot | % | % | arvot | % | % |
| -30 – -9 | 3.7 | 4.3 | 1 | 69 | 52 |
| -8 – 12 | 65 | 67 | 2 | 7 | 10 |
| 13 – 33 | 31 | 28 | 3 | 14 | 25 |
| | | | 4 | 10 | 13 |
| <i>nop-raji</i> | | | <i>nak-460</i> | | |
| arvot | % | % | arvot | % | % |
| 30 – 60 | 41 | 34 | 0 – 33 | 40 | 37 |
| 61 – 90 | 44 | 47 | 34 – 66 | 46 | 47 |
| 91 – 120 | 16 | 19 | 67 – 100 | 13 | 16 |
| <i>vkpv</i> | | | <i>kk</i> | | |
| arvot | % | % | arvot | % | % |
| ma | 18 | 15 | 1, 2 | 18 | 18 |
| ti | 12 | 13 | 3, 4 | 7.8 | 8.8 |
| ke | 13 | 14 | 5, 6 | 16.2 | 15.2 |
| to | 14 | 14 | 7, 8 | 17.2 | 15.9 |
| pe | 16 | 16 | 9, 10 | 17.8 | 17.5 |
| la | 11 | 15 | 11, 12 | 23 | 24 |
| su | 16 | 13 | | | |

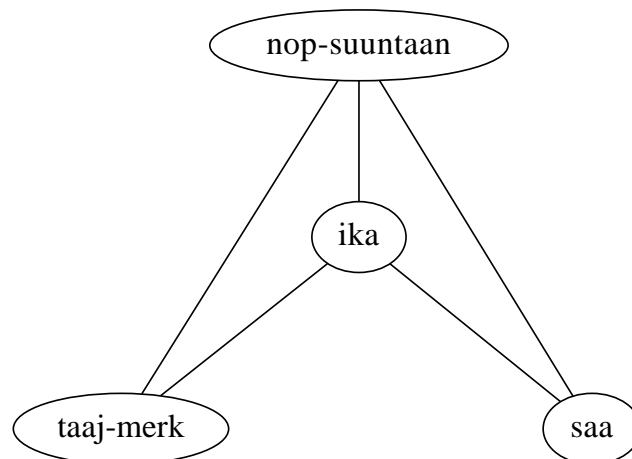
neonnettomuuksia muita ikäryhmiä enemmän taajamissa. Syynä tähän on todennäköisesti se, että yli 65-vuotiaat eivät aja yhtä paljon pitkiä matkoja kuin nuoremmat, vaan käyttävät autoa ennemminkin vain kauppa- ja muihin asiointimatkoihin. Nämä matkat voidaan yleensä suorittaa taajama-alueella, lähellä henkilön kotia.

Taulukosta 6.7 nähdään yli 65-vuotialle ja kaikenikäisille sattuneiden liikenneonnettomuuksien eroja keliolojen ja ajankohdan suhteen. Lämpötilan (*lampoila*), tien pinnan (*pinta*) ja kuukauden (*kk*) todennäköisyysjakaumien perusteella voidaan todeta, että vuodenajan ja sääolosuhteiden osalta ei juurikaan ole eroa. Kuukausien osalta huomataan, että vanhoille ihmisille sattuu kesällä hieman useammin liikenne-

neonnettumuksia kuin kaikenikäisille. Tämä voi johtua siitä, että monet vanhat ihmiset käyttävät autoaan vain sulan maan aikana. Lähes joka neljäs onnettomuus kummassakin ikäryhmässä sattuu marras-joulukuussa, kun on pimeää ja teillä mustaa jättä. Tarkasteltaessa viikonpäivien (*vkpv*) osalta todennäköisyysjakaumia huomataan, että lähes joka viides yli 65-vuotiaille sattunut liikenneonnettomuus on tapahtunut maanantaisin. Kun verrataan ikäihmisten jakaumaa kaikenikäisten jakaukseen, nähdään että ikäihmisille tapahtuu suhteellisesti enemmän liikenneonnettomuuksia maanantaisin ja sunnuntaisin, mutta vähemmän lauantaisin. Selvää syytä tähän ei ole.

Muuttujien *tunti* ja *valoisuus* (taulukko 6.7) jakaumista nähdään selvästi, että yli 65-vuotiaille sattuu liikenneonnettomuuksia pääasiassa päivällä ja valoisaan aikaan, kun taas koko väestön liikenneonnettomuudet jakautuvat melko tasaisesti ympäri vuorokauden yön tunteja lukuun ottamatta. Suurin osa yli 65-vuotiaista ihmisistä on eläkeläisiä, joten heillä on päivisin aikaa käydä asioimassa. Todennäköisesti eläkeläisiä on valoisaan aikaan huomattavasti enemmän liikenteessä kuin pimeään aikaan, joten myös liikenneonnettomuuksia heille sattuu enemmän päivällä.

Lopuksi tutkitaan vielä muutamaa kiinnostavaa muuttujaa yli 65-vuotiaitten liikenneonnettomuuksiin liittyen. Nämä muuttujat ovat *taaj-merk*, *nop-suuntaan* ja *saa*, ja niiden riippuvuussuhteet toisiinsa ja muuttujaan *ika* on esitetty kuvan 6.2 mallissa. Kuvan perusteella kaikilla kolmella edellä mainitulla muuttujalla on riippuvuussuhde muuttujaan *ika*. Riippuvuusmallissa on kuitenkin niin vähän muuttujia, ettei sen perusteella ole järkevää tehdä päätelmiä. Käytetään siis näiden muuttujien tutkimisessa taulukon 6.8 todennäköisyysjakaumia.



Kuva 6.2: Muuttujien *taaj-merk*, *nop-suuntaan* ja *saa* riippuvuussuhteet toisiinsa sekä muuttujaan *ika*.

Taulukosta 6.8 nähdään, että yli 65-vuotiaille sattuu kaikkiin ikäryhmiin verrattuna useammin liikenneonnettomuuksia taajamissa, kuten edellä esiteltujen tuloksien perusteella oletettiin. Myös ajonopeudet (*nop-suuntaan*) ovat vanhoilla ihmisillä

Taulukko 6.8: Muuttujien *nop-suuntaan*, *taaj-merk* ja *saa* arvojen todennäköisyysjakaumat muuttujan *ika* arvojen 65 – 93 suhteen ja muuttujien arvojen todennäköisyysjakaumat, kun muuttujalle *ika* ei ole valittu arvoa. Muuttujien arvojen selitykset ovat luvussa 4.

| Yli 65-vuotiaat | | | Yli 65-vuotiaat | | |
|---------------------|-----|-----|------------------|----|----|
| <i>nop-suuntaan</i> | | | <i>taaj-merk</i> | | |
| arvot | % | % | arvot | % | % |
| 30 – 60 | 39 | 33 | K | 26 | 19 |
| 61 – 90 | 35 | 32 | E | 74 | 81 |
| 91 – 120 | 26 | 36 | | | |
| <i>saa</i> | | | | | |
| arvot | % | % | | | |
| 1 | 35 | 34 | | | |
| 2 | 51 | 45 | | | |
| 3 | 1.0 | 2.1 | | | |
| 4 | 7.8 | 10 | | | |
| 5 | 3.5 | 6.6 | | | |
| 6,7 | 1.6 | 2.2 | | | |

painottuneet pienempiin nopeuksiin kun taas kaikenikäisillä ajonopeudet onnettomuustilanteissa ovat painottuneet suuriin nopeuksiin.

Aiemmin todettiin, että sääolosuhteiden osalta ei ole suuria eroja yli 65-vuotiaiden ja kaikenikäisten liikenneonnettomuuksissa. Muuttujan *saa* arvojen todennäköisyysjakaumista taulukossa 6.8 huomataan kuitenkin, että yli 65-vuotiaille tapahtuu kaikenikäisiin verrattuna hieman todennäköisemmin liikenneonnettomuuksia hyvän sään aikaan (arvot 1 ja 2, taulukko 4.1). Tähän varmasti vaikuttaa se, että yli 65-vuotiaat ovat pääosin eläkeläisiä, ja heidän ei tarvitse säästä riippumatta ajaa esimerkiksi työmatkoja. He voivat siis vapaammin valita, milloin ajavat autolla. Tämän takia todennäköisesti huonolla säällä (muuttujan *saa* arvot 3–7) yli 65-vuotiaita on vähemmän liikenteessä kuin hyvällä säällä.

7. YHTEENVETO

Tämän työn tarkoituksena oli selvittää, kuinka hyvin CosCo-tutkijaryhmän kehittämä B-course-ohjelmisto soveltuu liikenneonnettomuusaineiston tutkimiseen. Ohjelmistossa on kaksi tiedonlouhintamenetelmää. Toisella menetelmällä etsitään aineiston muuttujien välisiä riippuvuussuhteita ja toisella menetelmällä voidaan tutkia valitun muuttujan arvojen luokittelua muiden muuttujien avulla. Tässä työssä käytettiin vain riippuvuussuhteiden mallinnusmenetelmää. Tutkittavana aineistona työssä oli tietoja Pirkanmaan liikenneonnettomuuksista vuosilta 2004 – 2008. Käytetty aineisto on osa aineistoa, jossa on koko Suomessa samana aikana tapahtuneiden liikenneonnettomuuksien tiedot.

B-course-ohjelmiston riippuvuussuhteiden mallinnusmenetelmän toiminta perustuu bayesverkkoihin. Bayesverkot ovat verkkotodennäköisyysmalleja, jotka kuvaavat muuttujia ja niiden välisiä ehdollisia riippuvuuksia. Aineiston analysoinnin tuloksena menetelmä antaa parhaiten aineistoa kuvaavan riippuvuusmallin bayesverkon muodossa. Parhaan mallin ohjelmisto valitsee vertailemalla mallien todennäköisyyksiä keskenään. Tuloksena saadun riippuvuusmallin lisäksi ohjelmisto analysoi mallissa esiintyvien riippuvuussuhteiden vahvuudet; mitä vahvempia suhteet ovat, sitä todennäköisempi riippuvuusmalli on. Tämän lisäksi saadusta mallista annetaan kaksi erilaista kausaalisuusmallia. Toisessa kausaalisuusmallissa on huomioitu latentit muuttujat ja toisessa ei. Näiden tuloksena saatujen mallien ja interaktiivisen Java Playground -sovelluksen avulla käyttäjä voi analysoida ohjelmiston antamia tuloksia.

Aineiston analysoinnissa B-course-ohjelmistolla merkittävin rajoitus on tutkittavan aineiston koko. Koska ohjelmisto toimii internetissä ja on haluttu, että tulokset saadaan melko lyhyessä ajassa, ohjelmaan voidaan syöttää korkeintaan yhden megatavun kokoinen tiedosto. Tämän takia tässä työssä käytettiin aineistona vain osaa alkupeäisestä koko Suomen liikenneonnettomuusaineistosta.

Liikenneonnettomuusaineistoa tutkittaessa havaittiin, että jos tutkittavassa aineistossa on yli 15 muuttujaa, tuloksena saatu riippuvuusmalli on suuri ja melko vaikeasti tulkittavia. Tällaisessa tapauksessa aineiston muuttujat on hyvä jakaa pienempiin ryhmiin, joita voi tutkia erikseen, ja lopuksi yhdistää saadut tulokset tarkoitukseen sopivalla tavalla.

Liikenneonnettomuusaineiston muuttujien välisten kausaalisuussuhteiden tarkastelussa havaittiin, että B-course-ohjelmiston löytämien kausaalisuussuhteiden suunta oli joissakin tapauksissa päinvastainen kuin loogisin perustein pitäisi olla. Esimerkiksi ohjelmisto tarjoaa tien leveyden ja ajoratojen lukumäärän välille kausaalisuussuhdetta siten, että tien leveys on syy ja ajoratojen lukumäärä seuraus, tai kyseiselle kausaalisuussuhteelle on latentti syy. Loogisesti ajateltuna kuitenkin tien leveys on seurausta ajoratojen lukumäärästä, eikä päinvastoin. Ohjelmiston antamien kausaalisuussuhteiden suuntaan, kuten kaikkiin muihinkin tuloksiin, tulee siis suhtautua kriittisesti.

Tarkasteltaessa tuloksena saatua riippuvuusmallia Java Playground -sovelluksella huomataan, että jos muuttujalla on vähintään seitsemän numeerista arvoa, B-course-ohjelmisto luokittelee arvot kolmeen luokkaan, mikä voi joissakin tapauksissa hankaloittaa tulosten tarkastelua. Käyttäjä voi välttää ongelman jakamalla itse muuttujan arvot sopiviin luokkiin ennen aineiston syöttämistä ohjelmaan. Jos jokaista muuttujan arvoa halutaan tarkastella erikseen, numeeriset arvot voidaan korvata kirjaimilla tai sanoilla, jolloin B-course-ohjelmisto ei ryhmittele arvoja.

Tässä työssä tutkittiin Java Playground -sovelluksen avulla taajamissa ja taajamien ulkopuolella tapahtuneitten liikenneonnettomuuksien eroja, sekä eläinonnettomuuksia ja yli 65-vuotiaille sattuneita liikenneonnettomuuksia. Sovelluksella tarkasteltiin muiden muuttujien arvojen todennäköisyysjakaumien eroja, kun yhdelle muuttujalle valitaan tietty arvo.

Vuosien 2004 – 2008 aikana Pirkanmaalla sattuneista liikenneonnettomuuksista noin 85 % tapahtui taajamien ulkopuolella. Taajamien ulkopuolella sattuneissa onnettomuuksissa ajonopeudet ovat yleensä olleet suurempia kuin taajamissa, mikä johtuu siitä, että taajamissa ja taajamien ulkopuolella nopeusrajoitukset ovat erilaiset. Suuremmista ajonopeuksista huolimatta taajamien ulkopuolisissa liikenneonnettomuuksissa tulee harvemmin henkilö- tai muita omaisuusvahinkoja kuin taajamissa tapahtuneissa onnettomuuksissa.

Eläinonnettomuuksia Pirkanmaalla vuosien 2004 – 2008 aikana tapahtui noin 3200, joista noin 97 % oli hirvi- tai peuraonnettomuuksia. Lähes kaikki eläinonnettomuudet tapahtuivat taajamien ulkopuolella, joten ajonopeudet ovat yleensä olleet suuria. Eläinonnettomuudessa ei yleensä ole ollut muita osallisia kuin eläin ja siihen törmännyt ajoneuvo. Yli 70 % eläinonnettomuuksista tapahtuu hämärän tai pimeän aikaan, jolloin eläimet liikkuvat enemmän ja näkyvyys teillä on huonompi. Aineistossa ei ollut tietoa riista-aidoista, mutta eläinonnettomuuksien osalta olisi ollut mielenkiintoista tutkia riista-aidojen vaikutusta; tapahtuuko riista-aidatuilla tieosuuksilla vähemmän eläinonnettomuuksia kuin riista-aidattomilla tieosuuksilla.

Yli 65-vuotiaista kuljettajista johtuneita liikenneonnettomuuksia tapahtui Pirkanmaalla vuosien 2004 – 2008 aikana noin 600. Vanhemmille ihmisille sattui muu-

ta väestöä useammin onnettomuuksia taajamissa, ja ajonopeudet onnettomuuksissa olivat yleensä pienempiä. Noin 70 % yli 65-vuotiaiden liikenneonnettomuuksista tapahtui päivällä, valoisaan aikaan. Syynä tähän on todennäköisesti se, että vanhemmat ihmiset käyttävät autoa enimmäkseen asiointimatkoihin, jotka he voivat suorittaa päivällä, silloin kun työikäiset ihmiset ovat töissä. Jonkin verran muuta väestöä useammin yli 65-vuotiaiden ryhmässä onnettomuus on sattunut mieskuljettajalle. Tämä todennäköisesti johtuu siitä, että vanhemmissa ihmisissä naiset ajavat selvästi vähemmän autolla kuin miehet.

Työssä havaittiin, että B-course-ohjelmiston riippuvuuksien mallinnusmenetelmä soveltuu hyvin yksinkertaisten riippuvuussuhteiden löytämiseen käytetyn liikenneonnettomuusdatan tyyppisille aineistoille. Jos aineistossa on paljon muuttujia, se on syytä jakaa pienempiin osiin tuloksien tarkastelun helpottamiseksi. Tuloksina saatujen riippuvuus- ja kausaalisuusmallien muuttujien välisten suhteiden laatuun on kuitenkin hyvä suhtautua kriittisesti.

LÄHTEET

- [1] Äyrämö, S., Pirtala, P., Kauttonen, J., Naveed, K., Kärkkäinen T. Mining road traffic accidents. Jyväskylän yliopisto, 2009.
- [2] Järvenpää M., Turunen E. Suomessa 2004 – 2008 sattuneiden liikenneonnettomuuksien analysointia GUHA-tiedonlouhintamenetelmällä. Tampereen teknillinen yliopisto, 2012.
- [3] Grinstead, C. M., Snell J. L. Introduction to Probability. USA, 1997. 520 p.
- [4] Myllymäki P., Silander T., Tirri H., Uronen P. Bayesian Data Mining on the Web with B-Course. Proceedings IEEE International Conference on Data Mining, San Jose, USA, 29 November – 2 December 2001. pp. 626 – 629.
- [5] Myllymäki P., Silander T., Tirri H., Uronen P. B-Course: A Web Service for Bayesian Data Analysis. Proceedings of the 13th International Conference on Tools with Artificial Intelligence, Dallas, USA, 7 – 9 November, 2001. pp. 247 – 256.
- [6] B-course-ohjelmiston kotisivut: What kind of models are we actually talking about? [viitattu 20.4.2012] <http://b-course.cs.helsinki.fi/obc/models.html>
- [7] B-course-ohjelmiston kotisivut: Inferring causalities in the presence of latent variables. [viitattu 20.4.2012] http://b-course.cs.helsinki.fi/obc/not_naive_cause.html
- [8] Myllymäki P., Silander T., Tirri H., Uronen P. B-Course: A Web-based Tool for Bayesian and Causal Data Analysis. International Journal on Artificial Intelligence Tools 11(2002)3, pp. 369 – 387.
- [9] B-course-ohjelmiston kotisivut: How many models are there? [viitattu 23.4.2012] <http://b-course.cs.helsinki.fi/obc/howmanymodels.html>
- [10] Bernardo, J. M., Smith A. F. M. Bayesian Theory. Great Britain. 1995. 586 p.
- [11] Courant, R., John F. Introduction to Calculus and Analysis I. Germany. 1999. 661 p.
- [12] Box G. E. P., Tiao G. C. Bayesian Inference in Statistical Analysis. USA. 1973. 588 p.
- [13] Liikennevirasto, tierekisteri, tietosisällön kuvaus. Saatavilla: http://portal.liikennevirasto.fi/portal/page/portal/f/urakoitsijat_suunnittelijat/konsulteille/tierekisteri/tr_tietosisallon_kuvaus.pdf

A. LIIKENNEONNETTOMUUSTYYPIT

| Koodi | Selitys |
|----------|--|
| 0 | Samat ajosuunnat (mikään ajoneuvoista ei ollut kääntymässä) |
| 00 | ohitus |
| 01 | kaksoisohitus |
| 02 | kaistanvaihto oikealle |
| 03 | kaistanvaihto vasemmalle |
| 04 | kylkikosketus |
| 05 | ajo liikkeelle lähtevään ajoneuvoon |
| 06 | peräänajo jarruttavaan ajoneuvoon |
| 07 | muu peräänajo liikkuvaan ajoneuvoon |
| 08 | peräänajo liikenne-esteen takia pysähtyneeseen ajoneuvoon |
| 09 | muu onnettomuus |
| 1 | Samat ajosuunnat (jokin ajoneuvoista oli kääntymässä) |
| 10 | peräänajo käännyttäessä oikealle |
| 11 | muu törmäys käännyttäessä oikealle |
| 12 | peräänajo käännyttäessä vasemmalle |
| 13 | muu törmäys käännyttäessä vasemmalle |
| 14 | U-käännös samaan suuntaan kulkevan ajoneuvon eteen |
| 15 | pyöräilijä pyörätiellä, toinen ajoneuvo kääntyi oikealle |
| 16 | pyöräilijä pyörätiellä, toinen ajoneuvo kääntyi vasemmalle |
| 19 | muu onnettomuus |
| 2 | Vastakkaiset ajosuunnat (kohtaamisonnettomuus) |
| 20 | kohtaaminen suoralla |
| 21 | kohtaaminen kaarteessa |
| 22 | kohtaaminen ohitettaessa suoralla |
| 23 | kohtaaminen ohitettaessa kaarteessa |
| 24 | suistuminen väistämisen seurauksena |
| 29 | muu onnettomuus |

| Koodi | Selitys |
|----------|---|
| 3 | Vastakkaiset ajosuunnat (jokin ajoneuvoista oli kääntymässä) |
| 30 | kääntyminen vasemmalle vastaantulevan eteen tai kylkeen |
| 31 | kääntyminen samaan ajosuuntaan |
| 32 | kääntyminen eri ajosuuntiin |
| 33 | U-käännös vastaantulevan eteen |
| 34 | pyöräilijä pyörätiellä, vastaantuleva ajoneuvo kääntyi oikealle |
| 35 | pyöräilijä pyörätiellä, vastaantuleva ajoneuvo kääntyi vasemmalle |
| 36 | muu törmäys käännyttyessä oikealle |
| 39 | muu onnettomuus |
| 4 | Risteävät ajosuunnat |
| 40 | ajo risteäviä ajosuuntia suoraan |
| 41 | pyöräilijä pyörätiellä risteyksessä |
| 42 | pyöräilijä pyörätiellä muualla |
| 43 | junan ja ajoneuvon törmäys |
| 49 | muu onnettomuus |
| 5 | Risteävät ajosuunnat(jokin ajoneuvoista oli kääntymässä) |
| 50 | kääntyminen oikealle toisen eteen tai kylkeen |
| 51 | kääntyminen oikealle vastaantulevan eteen tai kylkeen |
| 52 | kääntyminen vasemmalle toisen eteen tai kylkeen |
| 53 | kääntyminen vasemmalle risteävän eteen tai kylkeen |
| 54 | yhtäaikainen vasemmalle kääntyminen |
| 55 | pyörätietä ajavan pyöräilijän kääntyminen ajoneuvon eteen tai kylkeen |
| 59 | muu onnettomuus |
| 6 | Jalankulkijaonnettomuus (suojatiellä) |
| 60 | jalankulkija suojatiellä ennen risteystä |
| 61 | jalankulkija suojatiellä risteyksen jälkeen |
| 62 | jalankulkija suojatiellä, ajoneuvo kääntyi vasemmalle |
| 63 | jalankulkija suojatiellä, ajoneuvo kääntyi oikealle |
| 64 | jalankulkija suojatiellä, suojatie risteyksen ulkopuolella |
| 65 | jalankulkija suojatiellä, suojatien eteen pysähtynyt auto |
| 69 | muu onnettomuus |

| Koodi | Selitys |
|----------|--|
| 7 | Jalankulkijaonnettomuus (muualla kuin suojatiellä) |
| 70 | jalankulkija tuli pysähtyneen ajoneuvon takaa |
| 71 | jalankulkija ylitti muutoin ajorataa suojatie ulkopuolella |
| 72 | jalankulkija pysähtyneenä ajoradalla |
| 73 | jalankulkija kulki liikenteen suuntaan |
| 74 | jalankulkija kulki liikennettä vastaan |
| 75 | jalankulkija jalkakäytävällä tai liikennekorokkeella |
| 76 | junan ja jalankulkijan törmäys |
| 79 | muu onnettomuus |
| 8 | Tieltä suistuminen |
| 80 | suistuminen oikealle suoralla |
| 81 | suistuminen vasemmalle suoralla |
| 82 | suistuminen oikealla oikealle kääntyvässä kaarteessa |
| 83 | suistuminen vasemmalle oikealle kääntyvässä kaarteessa |
| 84 | suistuminen oikealle vasemmalle kääntyvässä kaarteessa |
| 85 | suistuminen vasemmalle vasemmalle kääntyvässä kaarteessa |
| 86 | suistuminen tieltä risteyksessä |
| 89 | muu onnettomuus |
| 9 | Muu onnettomuus |
| 90 | eläinonnettomuus |
| 91 | törmäys oikeaan reunaan pysäköityyn ajoneuvoon |
| 92 | törmäys vasempaan reunaan pysäköityyn ajoneuvoon |
| 93 | törmäys liikennekorokkeeseen |
| 94 | törmäys esteeseen ajoradalla |
| 95 | kumoonajo ajoradalla |
| 96 | peruutusonnettomuus |
| 97 | matkustaja nousemassa tai poistumassa ajoneuvosta |
| 99 | muu onnettomuus |